

# PHONOLOGICAL AND PHONETIC COMPETENCE: BETWEEN GRAMMAR, SIGNAL PROCESSING, AND NEURAL ACTIVITY

EDITED BY: Ulrike Domahs, Hubert Truckenbrodt and Richard Wiese  
PUBLISHED IN: Frontiers in Psychology



# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2016 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-809-2

DOI 10.3389/978-2-88919-809-2

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# PHONOLOGICAL AND PHONETIC COMPETENCE: BETWEEN GRAMMAR, SIGNAL PROCESSING, AND NEURAL ACTIVITY

Topic Editors:

**Ulrike Domahs**, Free University of Bozen-Bolzano, Italy

**Hubert Truckenbrodt**, Centre for General Linguistics (ZAS), Germany

**Richard Wiese**, Philipps-Universität Marburg, Germany

The present collection of articles brings together experimental work in the field of segmental and prosodic processing and representation in phonology and phonetics. Contributions focus on the exploration of human cognitive, articulatory, and perceptual abilities dealing with all types of phonetic and phonological entities.

Main topics of investigation include: (1) sounds and sound-changing processes—systemic and functional aspects, (2) prosodic units such as syllables and metrical feet—systemic properties, processing, and phonetic consequences, and (3) tones as building blocks of the sentence melody—their relation to the level of linguistic expressions on the one hand, their phonetic realization (e.g., tonal height and contours) and perception on the other hand. In addition, topics (1) and (2) extend to the question how phonological representations are stored in the mental lexicon: specified minimally in terms of categorical phonological information or as variable phonetic imprint of the exemplars in the input.

Diagonally to these thematic domains, the present Research Topic shows a strong focus on up-to-date experimental approaches, going far beyond traditional linguistic analysis, and making use of psycho- and neurolinguistic methodologies.

**Citation:** Domahs, U., Truckenbrodt, H., Wiese, R., eds. (2016). Phonological and Phonetic Competence: Between Grammar, Signal Processing, and Neural Activity. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-809-2

# Table of Contents

- 05 Editorial: Phonological and Phonetic Competence: Between Grammar, Signal Processing, and Neural Activity**  
Ulrike Domahs, Hubert Truckenbrodt and Richard Wiese
- 08 Functional lateralization of lexical stress representation: a systematic review of patient data**  
Katja Häuser and Frank Domahs
- 12 Processing word prosody—behavioral and neuroimaging evidence for heterogeneous performance in a language with variable stress**  
Miriam Heisterueber, Elise Klein, Klaus Willmes, Stefan Heim and Frank Domahs
- 26 Phonetic detail in German syllable pronunciation: influences of prosody and grammar**  
Barbara Samlowski, Bernd Möbius and Petra Wagner
- 36 Use what you can: storage, abstraction processes, and perceptual adjustments help listeners recognize reduced forms**  
Katja Poellmann, Holger Mitterer and James M. McQueen
- 53 Processing of syllable stress is functionally different from phoneme processing and does not profit from literacy acquisition**  
Ulrike Schild, Angelika B. C. Becker and Claudia K. Friedrich
- 65 The direction of word stress processing in German: evidence from a working memory paradigm**  
Frank Domahs, Marion Grande, Walter Huber and Ulrike Domahs
- 73 Developing biases**  
Ruben van de Vijver and Dinah Baer-Henney
- 81 Do listeners recover “deleted” final /t/ in German?**  
Frank Zimmerer and Henning Reetz
- 91 An analysis of post-vocalic /s-ʃ/ neutralization in Augsburg German: evidence for a gradient sound change**  
Véronique Bukmaier, Jonathan Harrington and Felicitas Kleber
- 103 The role of predictability and structure in word stress processing: an ERP study on Cairene Arabic and a cross-linguistic comparison**  
Ulrike Domahs, Johannes A. Knaus, Heba El Shanawany and Richard Wiese
- 121 Evidence for the role of German final devoicing in pre-attentive speech processing: a mismatch negativity study**  
Hubert Truckenbrodt, Johanna Steinberg, Thomas K. Jacobsen and Thomas Jacobsen



**132** *Do syllables play a role in German speech perception? Behavioral and electrophysiological data from primed lexical decision*

Heidrun Bien, Jens Bölte and Pienie Zwitserlood

**144** *Production and perception of contrast: The case of the rise-fall contour in German*

Frank Kügler and Anja Gollrad



# Editorial: Phonological and Phonetic Competence: Between Grammar, Signal Processing, and Neural Activity

Ulrike Domahs<sup>1\*</sup>, Hubert Truckenbrodt<sup>2</sup> and Richard Wiese<sup>3</sup>

<sup>1</sup> Faculty of Education, Free University of Bozen-Bolzano, Bolzano, Italy, <sup>2</sup> Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany, <sup>3</sup> Institut für Germanistische Sprachwissenschaft, University of Marburg, Marburg, Germany

**Keywords:** prosody, phonetics, phonology, speech perception, speech production, language development, language change

## The Editorial on the Research Topic

## Phonological and Phonetic Competence: Between Grammar, Signal Processing, and Neural Activity

## INTRODUCTION

The present collection addresses the place and role of phonology (as an object of study, not as a scientific field) within a wider range of neighboring domains. Generally, the relevance of phonological structure in language may be claimed to derive from the fact that phonology constitutes a domain of its own within language (along with syntax, semantics, morphology), but also interfaces intimately with other domains such as cognition, articulation, and perception in general. From this dual nature, it follows that phonology may be an object of linguistic description and theory (for an overview see Goldsmith, 1995; de Lacy, 2012) as well as an object of cognitive and behavioral studies (for an overview see Cohn et al., 2012). Ideally, however, theoretical and empirical studies keep this dual nature of phonology in mind and pay attention to both sides of the coin.

Articles in the present Research Topic attempt to capture different aspects of this overall discussion. The starting point for this Research Topic was a Priority Programme on experimental research in phonology and phonetics funded by the German Science Foundation (DFG; SPP 1234). Based on this programme, the aim of this Research Topic is to draw together empirical work in the field of segmental and prosodic processing and representation and phonological theory.

Contributions address the interface of the speech sound systems investigated in phonology, the representations of articulated speech, perception, acquisition and processing established in phonetics, psycholinguistics, and neurolinguistics. Main topics of investigation include: (1) sounds and sound-changing processes—systemic and functional aspects, (2) prosodic units such as syllables and metrical feet—systemic properties, processing, and phonetic consequences, and (3) tones as building blocks of the sentence melody—their relation to the level of linguistic expressions on the one hand, their phonetic realization (e.g., tonal height and contours) and perception on the other hand. In addition, topics (1) and (2) extend to the question how phonological representations are stored in the mental lexicon: specified minimally in terms of categorical phonological information or as variable phonetic imprint of the exemplars in the input.

Diagonally to these thematic domains, the present Research Topic shows a strong focus on up-to-date experimental methods. Contributions go far beyond traditional linguistic analysis, and make use of psycho- and neuro-linguistic methods.

## OPEN ACCESS

### Edited and reviewed by:

Manuel Carreiras,  
Basque Center on Cognition, Brain  
and Language, Spain

### \*Correspondence:

Ulrike Domahs  
ulrike.domahs@unibz.it

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 21 November 2015

**Accepted:** 24 November 2015

**Published:** 23 December 2015

### Citation:

Domahs U, Truckenbrodt H and  
Wiese R (2015) Editorial: Phonological  
and Phonetic Competence: Between  
Grammar, Signal Processing, and  
Neural Activity.  
Front. Psychol. 6:1899.  
doi: 10.3389/fpsyg.2015.01899

## THE CONTRIBUTIONS

Sound and sound-changing processes are investigated by Bukmaier and colleagues, Schild and colleagues, Truckenbrodt and colleagues, van der Vijver and Baer, Poellmann and colleagues, and Zimmerer and Reetz. Bukmaier and colleagues present production and perception experiments that provide evidence for a process of sound change in which the neutralization of /s/ and /ʃ/ to /f/ before stops in Augsburg German is influenced by the Standard German contrast between /s/ and /ʃ/. The continuous function of the exposition to Standard German supports models that adhere to the exemplar theory of speech. Changes in sound perception during early childhood were studied by Schild and colleagues. In an ERP study with pre-schoolers, beginning readers, and adults, the authors investigated how stressed and unstressed syllables prime German word targets when prime and target overlap in phonemes and stress patterns. Age-related differences show that the processing of phonemes, but not the processing of stress is modulated by literacy acquisition.

An MMN-study devoted to investigate whether pre-attentive processing is sensitive to a syllable-related phonological process of German, namely final devoicing, was conducted by Truckenbrodt and colleagues. The authors found MMN effects for deviants violating final devoicing showing that even early pre-attentive auditory processing is modulated by syllable-related and automatic lexical phenomena. Final devoicing as the cause of voicing alternations in singular-plural pairs is also in the focus of the contribution by van de Vijver and Baer-Henney. In their production study, 5 and 7 years old children and adults produced plural forms out of pseudowords that required either voicing or vowel alternations. Age-related decrease of voicing and increase of vowel alternations show that generalizations are lexicon-based and rely on the frequencies of certain processes that vary between child and adult lexicon.

More indirectly connected to the topic of sound changing processes are two contributions on the production and perception of reduced forms displaying either sound deletions or reductions of phonological features. Poellmann and colleagues performed a series of eye-tracking experiments on the perception of reduced forms in which segments were either reduced or deleted. The experience with inconsistent pronunciations leads to a greater perceptual flexibility in dealing with other forms of reduction than does the experience with consistent pronunciations. The processing of reduced forms is also investigated by Zimmerer and Reetz. More specifically, they were interested in the sensitivity to compensatory acoustic cues left when a final /t/ is deleted, and investigated whether German listeners are able to reconstruct a final /t/ when confronted with reduced forms. They found that /t/ was reconstructed in only 45% of items presented. This finding is discussed in the light of the experimental methodology and stimuli used and the acoustic cues indexing final /t/ deletion in German.

The role of prosodic entities and/or their representation is investigated by Bien and colleagues, Samlowski and colleagues, Domahs and colleagues, Domahs and colleagues, Häuser and Domahs, Heisterueber and colleagues, and also

Schild and colleagues. In an ERP study using a word fragment priming paradigm, Bien and colleagues found effects that underpin the relevance of the syllable for language processing and lexical access. Samlowski and colleagues investigated the role of a number of prosodic and grammatical factors for syllable pronunciation in German. Some of these factors (word stress and sentence boundaries, lexical classes) were demonstrated to influence phonetic details (especially duration) of syllables corresponding to prefixes and function words.

How syllables are parsed into feet and whether feet are constructed beginning from the right or left edge of words has been investigated by Domahs and colleagues. The selection of the antepenultimate or final syllable as syllable bearing main stress in trisyllabic pseudowords is found to correlate with the working memory capacity of participants in a pseudoword production task.

A study on foot properties is presented by Domahs and colleagues. Their EEG results support evidence for bimoraic trochaic feet as processing units in the word stress system of Cairene Arabic. In addition, prosodic structure in Cairene Arabic is shown to be generated and constructed actively in online processing. The highly predictable word stress system does not lead to limitations in the sensitivity to word stress, i.e., there is no stress-deafness as predicted, among others, by Peperkamp and Dupoux (2002).

Regarding the question where lexical stress representations are functionally localized, Häuser and Domahs reviewed a series of published patient studies: all patients with a representational deficit in word stress processing had lesions in their language-dominant hemisphere. Word stress processing relies mainly on the functioning of the left hemisphere. However, Heisterueber and colleagues show that stress processing is also subject to inter-individual differences, as shown in an fMRI-study performed with German native speakers who participated in a sequence recall task testing the capacity to represent segmental and suprasegmental information on an abstract level. The authors report inter-individual differences in behavioral and neural activation patterns for word stress processing modulated by individual auditory processing and working memory capacities.

Finally, Kügler and Gollrad presented production and perception studies on contrastive meaning components of a rise-fall contour in German: a pitch accent carrying a particular meaning has a preference to occur with a context that triggers this particular meaning. Their findings suggest that the alignment and scaling of the accentual peak are sufficient to license a contrastive interpretation of the nuclear rise-fall contour.

## AUTHOR CONTRIBUTIONS

UD, HT, and RW conceived the topic of this Research Topic. All authors participated in the editorial work.

## ACKNOWLEDGMENTS

We would like to thank the German Science Foundation (Deutsche Forschungsgemeinschaft) for generously supporting the work presented in this Research Topic.

## REFERENCES

- Cohn, A. C., Fougeron, C., and Huffman, M. K. (eds.) (2012). *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press.
- de Lacy, P. (2012). *Cambridge Handbook of Phonological Theory*. Cambridge: Cambridge University Press.
- Goldsmith, J. A. (ed.) (1995). *The Handbook of Phonological Theory*. Cambridge, MA: Cambridge University Press.
- Peperkamp, S., and Dupoux, E. (2002). "A typological study of stress 'deafness'" in *Laboratory Phonology*, Vol., 7, eds C. Gussenhoven and N. Warner (Berlin: Mouton de Gruyter), 203–240.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Domahs, Truckenbrodt and Wiese. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Functional lateralization of lexical stress representation: a systematic review of patient data

Katja Häuser<sup>1,2</sup> and Frank Domahs<sup>3\*</sup>

<sup>1</sup> Department for Communication Sciences and Disorders, McGill University, Montreal, QC, Canada

<sup>2</sup> Centre for Research on Brain, Language and Music, McGill University, Montreal, QC, Canada

<sup>3</sup> Institute of Germanic Linguistics, Philipps-University Marburg, Marburg, Germany

## Edited by:

Richard Wiese, Philipps-Universität Marburg, Germany

## Reviewed by:

Marie Klopfenstein, Southern Illinois University Edwardsville, USA

Christina Samuelsson, Linköping University, Sweden

## \*Correspondence:

Frank Domahs, Institute of Germanic Linguistics, Philipps-University Marburg, Wilhelm-Röpke-Str., 6a, D-35032 Marburg, Germany  
e-mail: domahs@uni-marburg.de

According to the functional lateralization hypothesis (FLH) the lateralization of speech prosody depends both on its function (linguistic = left, emotional = right) and on the size of the units it operates on (small = left, large = right). In consequence, according to the FLH, lexical stress should be processed by the left (language-dominant) hemisphere, given its linguistic function and small unit size. We performed an exhaustive search for case studies of patients with acquired dysprosody due to unilateral brain damage. In contrast to previous reviews we only regarded dysprosody at the lexical level (excluding phrasal stress). Moreover, we focused on the representational stage of lexical stress processing, excluding more peripheral perceptual or motor deficits. Applying these criteria, we included nine studies reporting on 11 patients. All of these patients showed representational deficits in word stress processing following a lesion in their language-dominant hemisphere. In 9 out of 11 patients, it was the left hemisphere which was affected. This is a much more consistent pattern as found in previous reviews, in which less rigorous inclusion criteria may have blurred the pattern of results. We conclude that the representation of lexical stress crucially relies on the functioning of the language-dominant (mostly left) hemisphere.

**Keywords:** word stress, representational knowledge, left hemisphere, right hemisphere, acquired disorders of language, dysprosody

## INTRODUCTION

According to the functional lateralization hypothesis (FLH; Van Lancker, 1980; Van Lancker Sidtis et al., 2006) the lateralization of speech prosody depends both on its function and the size of the linguistic unit it operates on. Processing of prosody with an emotional function is assumed to be accomplished by the right hemisphere, whereas prosody with a linguistic function should be processed by the left or language-dominant hemisphere. Moreover, the right hemisphere is assumed to operate on larger scale linguistic units such as phrases or sentences, while small units such as syllables should be processed by the left (language dominant) hemisphere. In consequence, according to the FLH, lexical stress should be processed by the left (language-dominant) hemisphere, given its linguistic function and small unit size (Van Lancker, 1980; Wong, 2002).

One relevant source of evidence for the FLH are neuropsychological case studies. If lexical stress processing is found to be impaired in subjects with unilateral brain damage, this would provide insights into the neural substrates that are necessarily involved in the processing of this aspect of prosody. However, so far such studies have yielded mixed results with respect to the FLH, and different reviews have arrived at conflicting results (Baum and Pell, 1999; Wong, 2002). Whereas the authors in one review concluded that there is sufficient evidence in favor of a consistent involvement of left hemisphere substrates in lexical stress processing (Baum and Pell, 1999), another review found the results too inconclusive to fully support the hypothesis of

functional lateralization (Wong, 2002). These contradicting conclusions can partly be attributed to diverging methods and interpretations of the results. For example, Wong (2002) stated that since not all reviewed studies consistently include an LHD, RHD, and normal control group, some results are impossible to evaluate against the hypothesis of functional lateralization. Another potential problem is the fact that most previous studies have intermixed tasks involving different stages of lexical stress processing (such as perception, representation, and production), although it seems implausible that these processing stages are accomplished by the same neural regions at all (for a review, see Zatorre and Gandour, 2008). This has possible consequences for lateralization according to the FLH and could also explain why previous reviews did not reach a consistent conclusion in this matter. Finally, existing reviews often included clinical case studies conducted in English, some of which have insufficiently distinguished the size of the linguistic units under consideration. In some studies, compound noun phrases (green 'house vs. 'greenhouse) have been investigated on the same level as noun/verb minimal pairs ('convict vs. con'vict). Such an approach is potentially problematic, since noun phrases have greater semantic and syntactic complexity than compound nouns or simplex nouns and verbs (Wasow, 1997). Consequently, 'green house is not minimally distinct from green 'house in regards to word stress alone. Crucially, they also differ in the size of linguistic units involved which has implications for the lateralization of processing according to the FLH. In sum, various reasons ranging from differing methodologies

to contrasting interpretations could explain the rather mixed evidence that has been discussed with respect to the FLH so far.

The goal of the present study was to review existing case reports with respect to the functional lateralization of lexical stress. In contrast to previous reviews (Baum and Pell, 1999; Wong, 2002), we only considered clinical case studies that investigated prosody at a purely lexical level, thus excluding studies on noun phrase or compound noun stress processing. Moreover, we focused on the representational stage of lexical stress processing, excluding perceptual and articulatory deficits. Our aim was to evaluate evidence informative to the claim that lexical stress is represented in the language-dominant (i.e., mostly left) hemisphere.

## METHODS

We conducted an exhaustive search on the data bases Google Scholar, PubMed, and Entrez, using the search terms *lexical stress AND brain damage*, *lexical stress AND hemisphere*, and *lexical stress AND aphasia*. We discarded all studies that did not focus on individuals with unilateral brain damage and/or did not investigate stress assignment at a purely lexical level (for example, studies on tone or phrase level stress). In addition, we excluded all studies in which prosodic impairments in speech production could also result from more peripheral perceptual or articulatory difficulties (e.g., cases of dysarthric or apraxic impairment). After the application of these exclusion criteria, 12 articles remained for analysis, reporting on 15 patients with representational impairments in lexical stress processing (meaning that these patients displayed impairments in stress assignment). We reviewed all 12 studies for hemispheric site of lesion and language-dominant hemisphere of the patient. If both types of information were missing and could not be inferred based on the information provided

by the authors, the study and/or subject was excluded from further analysis, resulting in the exclusion of four studies/subjects (excluded studies: Lloyd, 1999; Howard and Smith, 2002; Janssen, 2003. excluded subject: DE in Black and Byng, 1986). An overview on all studies and patients included is provided in **Table 1**.

The languages spoken by the patients included in our analyses were English, German, and Italian—all languages with variable stress. This means that although in all three languages, word stress assignment shows some regularities (for an overview, see Van der Hulst, 1999), the assignment of stress to individual words cannot be inferred by phonemic or orthographic rules alone and thus requires activation of word-specific (i.e., lexical) phonological representations (Miceli and Caramazza, 1993). The word stress errors reported by the studies included in this review particularly affected words with infrequent or “irregular” stress patterns (Coltheart et al., 1983; Chiacchio et al., 1993; Miceli and Caramazza, 1993; Cappa et al., 1997; Rozzini et al., 1997; Galante et al., 2000; Laganaro et al., 2002; Janssen and Domahs, 2008), typically leading to shifts in stress assignment to the most frequent pattern (“over-regularisations”, e.g., Marshall and Newcombe, 1973; Black and Byng, 1986; Cappa et al., 1997; Laganaro et al., 2002; Janssen and Domahs, 2008).

## RESULTS

### CLASSIFICATION BASED ON HANDEDNESS

This analysis is based on the fact that handedness is closely related to hemispheric dominance for language (e.g., Knecht et al., 2000). Out of the ten studies with 12 patients that remained in the pool (see **Table 1**), data on both the impaired hemisphere and the patient’s handedness were available in eight cases. All of these eight patients presented with systematic errors in stress assignment following a lesion in their language-dominant hemisphere

**Table 1 | Table of patients included in the review.**

Study	Patient	Hand	Hemisphere	Site	Syndrome	Fluency	Etiology	Nam (%)	Read (%)	Repet (%)	Lex. Dec. (%)
Chiacchio et al., 1993	CA	R	LH	T	Surface dyslexia, PPA	FL	DEG		30		
Galante et al., 2000	RM	R	LH	T, Ven	Surface dyslexia, PPA	FL	DEG		21		
Laganaro et al., 2002	MS	L + R	LH	x	Aphasia, Apraxia	NFL	CVA	10	12	10	7
Miceli and Caramazza, 1993	CLB	R	LH	T	Surface dyslexia	FL	CVA		26		
Rozzini et al., 1997	AC	R	LH	T, Ven	Surface dyslexia	FL	DEG		27		
Janssen and Domahs, 2008	HAT	R	LH, RH	T, Sub	Surface dyslexia, PPA	FL	DEG		25	2	
Marshall and Newcombe, 1973	ST	x	LH	T, P	Surface dyslexia	FL	OHI		xs		
	JC	x	LH	T, P	Surface dyslexia	FL	OHI		x		
Black and Byng, 1986	RW	x	LH	T, P	Deep dyslexia	FL	OHI		3		
	HRM	x	LH	x	Deep dyslexia, Agrammatism	NFL	CVA		3		
Coltheart et al., 1983	AB	L	RH	F	Surface dyslexia	x	OHI		x		
Cappa et al., 1997	GM	R	LH	T	Conduction aphasia	FL	OHI	18	14		

Hand, reported handedness; Hemisphere, lesioned hemisphere; Site, site of lesion (if available); T, Temporal; Ven, Ventricle; P, Parietal; F, Frontal.

Etiology: DEG, Degenerative; OHI, Open Head Injury; CVA, Cerebral Vascular Accident.

Tasks: nam, picture naming; read, reading; repet, repetition; lex. dec., lexical decision task performance is indicated in % errors. x's indicate missing information.



(which was the LH in seven patients and the RH in one patient), as inferred from handedness.

### CLASSIFICATION BASED ON LINGUISTIC IMPAIRMENT

All 12 patients (including the four cases where handedness information was not available) showed major linguistic impairments, suggesting that they suffered from lesions in their language-dominant hemisphere. This yields a total of 12 out of 12 patients who showed representational deficits in word stress processing following a lesion in their language-dominant hemisphere. In 10 out of 12 patients, it was the left hemisphere which was affected.

### DISCUSSION

The goal of this study was to review evidence from acquired language impairment regarding the functional lateralization hypothesis (FLH, Van Lancker, 1980; Van Lancker Sidtis et al., 2006), which states that function and size of a prosodic unit determine the cortical hemisphere it is processed in. Specifically, we were interested in the representation of lexical stress, which according to the FLH is a property of the left, language-dominant hemisphere. To this end, we reviewed clinical case studies that focused on brain-damaged patients with representational impairments in lexical stress assignment. Ten studies reporting on 12 cases remained for analysis after the application of all exclusion criteria. The results showed that in all of these patients, impairments in lexical stress assignment followed a lesion in the language-dominant hemisphere. In contrast to earlier reviews, which have arrived at mixed results, our data thus fully support the functional lateralization hypothesis.

The sample of studies that met our inclusion criteria was rather small, given that only few studies addressed the representation (rather than perception or articulation) of lexical stress. However, our rigorous and hypothesis-driven approach yielded a very clear pattern of results, in comparison to previous reviews that investigated speech prosody. In fact, a closer look at patients which were excluded from our analyses because they did not fulfill our criterion of a representational impairment (either because of a speech-motor deficit, e.g., dysarthria, or because of a non-specified deficit affecting lexical stress processing) revealed a much less consistent pattern: 82 of those patients had a left-hemisphere lesion, in comparison to 74 patients with right-hemisphere damage. Furthermore, 65 patients were reported to have lesions at the side of their dominant hand. Clearly, allowing for less precision in the nature of lexical dysprosody (as in previous reviews) would have led to a more impressive number of cases but to a blurred pattern of results. After all, it is highly plausible that perceptual and motor stages of lexical stress processing are subserved by bilateral brain areas whereas the more abstract linguistic representation of word prosody may reside in the language-dominant (left) hemisphere. This could explain the mixed evidence that earlier reviews yielded with respect to the FLH (Baum and Pell, 1999; Wong, 2002).

Our findings are consistent with evidence from dichotic listening showing that stress typicality effects (indicative of the representational stage of stress processing) only appeared in repetition and noun/verb-classification when stimuli were presented to the right ear/left hemisphere (Arciuli and Slowiczek, 2007).

More generally, our findings are also consistent with previous studies (Baum and Pell, 1999) that have rejected a strict division of labor regarding the hemispheric representation of prosody. Even though our results support the notion that lexical stress is a property of the language-dominant hemisphere, it seems that any global “all-left” or “all-right” account with respect to the hemispheric lateralization of *all* prosodic functions is an oversimplification and fails to account for the data. In this context, it seems that to date the FLH is the most promising account put forward to describe the neural substrates of prosody, since it does not set up an all-or-none division for prosodic functions but allows for gradedness of prosodic representation, depending on their function and the size of the processing units involved. This claim is also substantiated by findings in neuro-imaging, which have demonstrated bilateral cortical activations for lexical stress processing (Aleman et al., 2005; Wildgruber et al., 2006; Klein et al., 2011; Domahs et al., 2013). Yet, it is the methodological strength of lesion studies to highlight the functional relevance of brain regions for cognitive functions (Rorden and Karnath, 2004).

In sum, based on the data at hand we conclude that the representation of lexical stress crucially relies on the functioning of the language-dominant (mostly left) hemisphere.

### ACKNOWLEDGMENTS

This research was supported by a grant from the LOEWE initiative of excellence of the Hessian Ministry of Research and the Arts (project LingBas).

### REFERENCES

- Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Haan, E. H., and Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cereb. Cortex* 15, 221–228. doi: 10.1093/cercor/bbh124
- Arciuli, J., and Slowiczek, L. M. (2007). The where and when of linguistic word-level prosody. *Neuropsychologia* 45, 2638–2642. doi: 10.1016/j.neuropsychologia.2007.03.010
- Baum, S. R., and Pell, M. D. (1999). The neural bases of prosody: insights from lesion studies and neuroimaging. *Aphasiology* 13, 581–608. doi: 10.1080/026870399401957
- Black, M., and Byng, S. (1986). Prosodic constraints on lexical access in reading. *Cogn. Neuropsychol.* 3, 369–409. doi: 10.1080/02643298608252028
- Cappa, S. F., Nespore, M., Ielasi, W., and Miozzo, A. (1997). The representation of stress: evidence from an aphasic patient. *Cognition* 65, 1–13. doi: 10.1016/S0010-0277(97)00024-3
- Chiacchio, L., Grossi, D., Stanzione, M., and Trojano, L. (1993). Slowly progressive aphasia associated with surface dyslexia. *Cortex* 29, 145–152. doi: 10.1016/S0010-9452(13)80219-5
- Coltheart, M., Masterson, J., Byng, S., Prior, M., and Riddoch, J. (1983). Surface dyslexia. *Q. J. Exp. Psychol.* 35A, 469–495. doi: 10.1080/14640748308402483
- Domahs, U., Klein, E., Huber, W., and Domahs, F. (2013). Good, bad and ugly word stress – fMRI evidence for foot structure driven processing of prosodic violations. *Brain Lang.* 125, 272–282. doi: 10.1016/j.bandl.2013.02.012
- Galante, E., Tralli, A., Zuffi, M., and Avanzi, S. (2000). Primary progressive aphasia: a patient with stress assignment impairment in reading aloud. *Neurol. Sci.* 21, 39–48. doi: 10.1007/s100720070117
- Howard, D., and Smith, K. (2002). The effects of lexical stress in aphasic word production. *Aphasiology* 16, 198–237. doi: 10.1080/02687040143000546
- Janssen, U. (2003). Stress assignment in German patients with surface dyslexia. *Brain Lang.* 87, 114–115. doi: 10.1016/S0093-934X(03)00225-6
- Janssen, U., and Domahs, F. (2008). Going on with optimized feet: evidence for the interactions between segmental and metrical structure in phonological encoding from a case of primary progressive aphasia. *Aphasiology* 22, 1157–1175. doi: 10.1080/02687030701820436

- Klein, E., Domahs, U., Grande, M., and Domahs, F. (2011). Neuro-cognitive foundations of word stress processing - evidence from fMRI. *Behav. Brain Funct.* 7, 15–32. doi: 10.1186/1744-9081-7-15
- Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., et al. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain* 123, 2512–2518. doi: 10.1093/brain/123.12.2512
- Laganaro, M., Vacheresse, F., and Frauenfelder, U. H. (2002). Selective impairment of lexical stress assignment in an Italian-speaking aphasic patient. *Brain Lang.* 81, 601–609. doi: 10.1006/brln.2001.2550
- Lloyd, A. J. (1999). Comprehension of prosody in Parkinson's Disease. *Cortex* 35, 389–402. doi: 10.1016/S0010-9452(08)70807-4
- Marshall, J. C., and Newcombe, F. (1973). Patterns of paralexia: a psycholinguistic approach. *J. Psycholinguist. Res.* 2, 175–198. doi: 10.1007/BF01067101
- Miceli, G., and Caramazza, A. (1993). The assignment of word stress in oral reading: evidence from a case of acquired dyslexia. *Cogn. Neuropsychol.* 10, 273–296. doi: 10.1080/02643299308253465
- Rorden, C., and Karnath, H. O. (2004). Using human brain lesions to infer function: a relic from a past era in the fMRI age? *Nat. Rev. Neurosci.* 5, 813–819. doi: 10.1038/nrn1521
- Rozzini, L., Bianchetti, A., Lussignoli, G., Cappa, S., and Trabucchi, M. (1997). Surface dyslexia in an Italian patient with semantic dementia. *Neurocase* 3, 307–312. doi: 10.1080/13554799708405014
- Van der Hulst, H. (ed.). (1999). *Word Prosodic Systems in the Languages of Europe. Empirical Approaches to Language Typology*, Vol. 4. Berlin, New York, NY: de Gruyter.
- Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Papers Linguist. Int. J. Hum. Commun.* 13, 201–277. doi: 10.1080/08351818009370498
- Van Lancker Sidtis, D., Pachana, N., Cummings, J. L., and Sidtis, J. J. (2006). Dysprosodic speech following basal ganglia insult: toward a conceptual framework for the study of the cerebral representation of prosody. *Brain Lang.* 97, 135–153. doi: 10.1016/j.bandl.2005.09.001
- Wasow, T. (1997). Remarks on grammatical weight. *Lang. Variation Change* 9, 81–105. doi: 10.1017/S0954394500001800
- Wildgruber, D., Ackermann, H., Kreifelts, B., and Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog. Brain Res.* 156, 249–268. doi: 10.1016/S0079-6123(06)56013-3
- Wong, P. C. M. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Res. Bull.* 59, 83–95. doi: 10.1016/S0361-9230(02)00860-2
- Zatorre, R., and Gandour, J. T. (2008). Neural specialization for speech and pitch: moving beyond the dichotomies. *Philos. Trans. R. Soc. Biol.* 363, 1087–1104. doi: 10.1098/rstb.2007.2161

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 January 2014; accepted: 26 March 2014; published online: 10 April 2014.

Citation: Häuser K and Domahs F (2014) Functional lateralization of lexical stress representation: a systematic review of patient data. *Front. Psychol.* 5:317. doi: 10.3389/fpsyg.2014.00317

This article was submitted to Language Sciences, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Häuser and Domahs. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Processing word prosody—behavioral and neuroimaging evidence for heterogeneous performance in a language with variable stress

Miriam Heisterueber<sup>1,2\*</sup>, Elise Klein<sup>2,3,4</sup>, Klaus Willmes<sup>1,2,4</sup>, Stefan Heim<sup>1,5,6</sup> and Frank Domahs<sup>7</sup>

<sup>1</sup> Section Neurological Cognition Research, Department of Neurology, Uniklinik RWTH Aachen, Aachen, Germany

<sup>2</sup> Faculty of Medicine, Brain Imaging Facility of the Interdisciplinary Centre for Clinical Research, Uniklinik RWTH Aachen, Aachen, Germany

<sup>3</sup> KMRC – Knowledge Media Research Center, Tuebingen, Germany

<sup>4</sup> Section Neuropsychology, Department of Neurology, Uniklinik RWTH Aachen, Aachen, Germany

<sup>5</sup> Department of Psychiatry, Psychotherapy and Psychosomatics, Uniklinik RWTH Aachen, Aachen, Germany

<sup>6</sup> Research Centre Juelich, Institute of Neuroscience and Medicine (INM-1), Juelich, Germany

<sup>7</sup> Institute of Germanic Linguistics, Philipps University, Marburg, Germany

## Edited by:

Richard Wiese, Philipps-Universität Marburg, Germany

## Reviewed by:

Maren Schmidt-Kassow, Goethe University, Germany

Anna J. Simmonds, Imperial College London, UK

## \*Correspondence:

Miriam Heisterueber, Section Neurological Cognition Research, Department of Neurology, Uniklinik RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany  
e-mail: miriam.heisterueber@rwth-aachen.de

In the present behavioral and fMRI study, we investigated for the first time interindividual variability in word stress processing in a language with variable stress position (German) in order to identify behavioral predictors and neural correlates underlying these differences. It has been argued that speakers of languages with variable stress should perform relatively well in tasks tapping into the representation and processing of word stress, given that this is a relevant feature of their language. Nevertheless, in previous studies on word stress processing large degrees of interindividual variability have been observed but were ignored or left unexplained. Twenty-five native speakers of German performed a sequence recall task using both segmental and suprasegmental stimuli. In general, the suprasegmental condition activated a subcortico-cortico-cerebellar network including, amongst others, bilateral inferior frontal gyrus, insula, precuneus, cerebellum, the basal ganglia, pre-SMA and SMA, which has been suggested to be dedicated to the processing of temporal aspects of speech. However, substantial interindividual differences were observed. In particular, main effects of group were observed in the left middle temporal gyrus (below vs. above average performance in stress processing) and in the left precuneus (above vs. below average). Moreover, condition (segmental vs. suprasegmental) and group (above vs. below average) interacted in the right hippocampus and cerebellum. At the behavioral level, differences in word stress processing could be partly explained by individual performance in basic auditory perception including duration discrimination and by working memory performance (WM). We conclude that even in a language with variable stress, interindividual differences in behavioral performance and in the neuro-cognitive foundations of stress processing can be observed which may partly be traced back to individual basic auditory processing and WM performance.

**Keywords:** word stress, fMRI, interindividual differences, segmental processing, stress processing

## INTRODUCTION

In some languages (e.g., Czech, Finnish, Polish, Turkish, Persian, or French) main stress always falls on the same position within a word (fixed stress; for a typological overview see Van der Hulst, 1999). In those languages, no minimal pairs of words exist which do only differ in terms of their stress position. Accordingly, in fixed stress languages word stress is not contrastive and does not carry lexical information. In consequence, the processing and representation of word stress is not particularly relevant in the use of such languages. In this vein, it has been repeatedly reported that speakers of languages with fixed stress encounter difficulties when confronted with tasks requiring processing or representation of word prosody (Dupoux et al., 1997; Peperkamp et al., 1999, 2010; Mehler et al., 2004; Domahs et al., 2012, 2013a).

In contrast, other languages (e.g., English, Spanish, Russian, or German) have variable stress positions. Word stress may be contrastive, carrying lexical information. Thus, there may be minimal pairs, which only differ in their suprasegmental make-up, i.e., stress pattern, their segmental sequence being identical (e.g., German verbs *umfahren* vs. *úmfahren*, to drive around vs. to knock over). Therefore, the processing and representation of word stress is particularly relevant in languages with variable stress and speakers of those languages are typically found to be highly sensitive to suprasegmental manipulations, showing relatively good performance in a variety of tasks tapping on word stress (Domahs et al., 2008; Molczanow et al., 2013; for a direct comparison between speakers of a language with fixed stress (French) and with variable stress (Spanish or German) see Dupoux et al., 2001, 2008; Schmidt-Kassow et al., 2011a).

However, comparing speakers of different languages typically ignores the possibility that there may be substantial interindividual variability in stress processing performance even within a given language. Thus, the present study addresses the questions whether there are interindividual differences in stress processing in a language with variable stress (German) and, if so, which neural correlates may underlie those differences. Before the details of the present study will be outlined, a brief summary of research on stress processing will be given by describing word stress assignment in German and discussing evidence on the neuronal basis of stress processing.

### WORD STRESS ASSIGNMENT IN GERMAN

Given that German is a language with variable stress, the stress pattern of individual words is largely unpredictable and has thus to be lexicalized (Eisenberg, 2006; Domahs et al., 2008). This lexical knowledge can be used to distinguish between the elements of minimal pairs and to activate the correct meaning related to each of the members of a minimal pair. Beyond complete lexicalization, there are some rules and regularities in German stress assignment which become apparent, when participants are asked to pronounce pseudowords or have to deal with stress violations:

- (a) Only one of the final three syllables of a word can bear main stress (“three syllable window,” Vennemann, 1990). Thus, words can have ultimate stress (U, final syllable stressed), penultimate stress (PU, prefinal syllable stressed), or antepenultimate stress (APU, semi-prefinal syllable stressed).
- (b) Stress assignment is influenced by syllabic structure, in particular by the syllable weight of the final syllables (Tappeiner et al., 2007; Domahs et al., 2008; Janssen and Domahs, 2008; Roettger et al., 2012) such that words with open final and/or closed pre-final syllables are predominantly stressed on the penultimate syllable. Complex final syllables typically lead to main stress on the final syllable. Antepenultimate stress is typically found, when the penult is open and the final syllable is closed.
- (c) Main stress may be conceived as surface expression of metrical foot structure (which is determined by syllable weight) such that strong feet bear main stress. As prosodic feet are typically binary (i.e., containing two syllables which form a trochee, Knaus and Domahs, 2009), but heavy final syllables are parsed into non-branching feet, ultimate and antepenultimate stress can be seen as structurally similar in contrast to penultimate stress (Domahs et al., 2008, 2013b; Haake et al., 2013).
- (d) Penultimate stress is the most frequent pattern in German. Féry (1998) found that 73% of German bisyllabic words are stressed on the penult. In this light, it has been debated whether in German the penultimate stress pattern can be regarded as the default (e.g., Eisenberg, 1991; Kaltenbacher, 1994; Wiese, 1996; Levelt et al., 1999) or not (Giegerich, 1985; Vennemann, 1991; Féry, 1998; Domahs et al., 2008; Janssen and Domahs, 2008; Roettger et al., 2012).

Phonetically, German word stress is marked by a combination of the following cues: duration, (global) intensity, fundamental

frequency (pitch), vowel formants and voice quality (for a comprehensive overview see Lintfert, 2010). Haake et al. (2013) found a significant relationship between auditory perception of duration cues and the representation of word stress both in children with specific language impairment and in typically developing children acquiring German. Heim and Alter (2006, 2007) provided EEG evidence that context stress, e.g., in a sentence, can be used as additional information to identify stress patterns.

### THE NEURAL BASES OF WORD STRESS PROCESSING

There are currently only very few functional neuroimaging studies investigating the neural correlates of word stress processing (Aleman et al., 2005; Klein et al., 2011; Domahs et al., 2013b). In the study by Aleman et al. (2005) participants had to identify weak-initial and strong-initial words. The bilateral supplementary motor area (SMA) and the left inferior frontal gyrus (IFG), the superior temporal gyrus (STG) as well as the superior temporal sulcus (STS), and the insula were associated with the processing of word stress compared to a semantic control condition. In the study by Klein et al. (2011) participants were asked to solve an identity matching task with pseudowords. Processing of word stress minimal pairs as compared to segmental minimal pairs was associated with activation in a bilateral fronto-temporal network. Klein et al. (2011) suggested that there is a basic system for word stress processing in the left hemisphere, whereas the right hemisphere supports the left in case of increasing task difficulty. Domahs et al. (2013b) investigated the neural correlates of processing correctly vs. incorrectly stressed words. They observed activations of the left posterior angular and retrosplenial cortex when contrasting the processing of correct vs. incorrect stress. In the inverse contrast, bilateral STG were found to be involved. The analysis of severe vs. mild stress violations revealed activations of the left superior temporal and left anterior angular gyrus. Frontal activations, including Broca’s area and its right homolog, were found when contrasting mild with severe stress violations.

With respect to interindividual differences in stress processing, Boecker et al. (1999) performed an ERP study using a word stress discrimination task. Based on the median split of the behavioral outcome, they defined two groups of participants: good and poor performers. The authors found a significant N400-effect for sequence-final words with a weak-strong pattern only in the group of good performers, but not in the group of poor performers, providing first evidence to the possibility of substantial interindividual differences in word stress processing in a language with variable stress (Dutch).

### THE PRESENT STUDY

While differences in word stress processing between speakers of languages with fixed vs. variable stress have been described repeatedly (Dupoux et al., 2001, 2008; Peperkamp et al., 2010; Schmidt-Kassow et al., 2011a,b), interindividual differences within one type of language—although observed—remained largely ignored or unexplained (Boecker et al., 1999; Peperkamp et al., 1999; Domahs et al., 2008, 2013b; Dupoux et al., 2010). In general, it has been argued that speakers of a language with variable stress should perform relatively well in word stress processing (Dupoux et al., 1997, 2001, 2008, 2010; Peperkamp et al.,

1999; Schmidt-Kassow et al., 2011a). Although interindividual variance in word stress processing in German has not been the focus of previous research, such variability has been observed (albeit ignored) in adult participants in previous studies (Domahs et al., 2008, 2013b). In a recent study, (Haake et al., 2013) reported interindividual variability in word stress processing in both children with specific language impairment and typically developing children. This variance was at least partly predicted by individual perceptual processing of auditory cues related to word stress (e.g., duration).

The aim of the current study was to investigate interindividual performance differences in the processing of word stress. To this end, native speakers of German had to perform a variant of a sequence recall task, adapted from Dupoux et al. (2001; see also Haake et al., 2013). Studies on languages with fixed stress using this task have shown that when demands on working memory increase, performance of speakers of such languages in reproducing pseudoword minimal pairs (e.g., *míkuta* vs. *mikutá*) decreases disproportionately (Dupoux et al., 1997, 2001). We used a suprasegmental variant of this task to investigate interindividual heterogeneity in word stress processing in native speakers of German, a language with variable stress, while a segmental variant of this task served as a control condition. Note that speakers of German should be highly familiar with both suprasegmental and segmental features since both are essential in the use of this language.

In sum, the research questions of the present study were the following: (i) Are there substantial interindividual differences in word stress processing within a group of native speakers of German, a language where this feature is functional? (ii) Which neural correlates in functional magnetic resonance imaging (fMRI) are associated with word stress processing in good and poor performers? Following the results of previous neuroimaging studies on word stress processing (Aleman et al., 2005; Klein et al., 2011; Domahs et al., 2013b), we expected to find clusters of activated voxels in the left IFG, the bilateral superior temporal gyrus/sulcus and in the insula as well as bi-hemispheric activation in the SMA. (iii) Can predictors for interindividual variability be identified (e.g., working memory abilities and/or basic auditory processing)?

## MATERIALS AND METHODS

### PARTICIPANTS

Twenty-five right-handed native German-speaking healthy volunteers (nine female; mean age = 28.8 years, *SD* = 10.1 years) participated in the study after having given their written informed consent. The study was approved by the Institutional Review Board of the Medical Faculty at RWTH Aachen University (EK 182/06).

### STIMULI

Stimulus material consisted of trisyllabic pseudowords obeying German phonotactic constraints. The pseudowords were built from five different consonants (plosives: p, t, k; nasals: n, m) and three different vowels (a, u, i). All items had the same syllable structure (CV.CV.CV). Minimal pairs of pseudowords were created such that they either differed only with respect to word stress

(suprasegmental condition, SSEG) or only with respect to one consonant (segmental condition, SEG). There were two suprasegmental contrasts and two segmental contrasts, each consisting of two items, respectively (see Table 1). In the suprasegmental condition, penultimate stress (PU) was compared to final stress (U) and antepenultimate stress (APU) was contrasted to final stress (U). In the segmental condition, the consonants differed either in place of articulation (POA) or in a combination of place and manner of articulation (MOA). In the POA condition the consonants /m/ vs. /n/ and /k/ vs. /p/ were contrasted, whereas in the MOA condition /t/ vs. /f/ and /k/ vs. /s/ were contrasted.

For each type of stimulus, different tokens were recorded such that in each minimal pair one token was spoken by a female speaker (native speaker of Polish) and one token was spoken by a male speaker (native speaker of Persian), with the order being counterbalanced across conditions. Each pseudoword was recorded multiple times from each speaker so that different tokens from the same word were presented in the experiment. In this way, phonetic variance of stimuli was increased, disfavoring purely auditory/phonetic strategies and encouraging a more abstract, phonological type of target comparison. The duration of the pseudowords was approximately 1000 ms. Stimuli were recorded using Amadeus Pro sound editing software (HairerSoft, Kenilworth, UK).

### PRETEST PROCEDURE

Each participant completed pretests to evaluate his/her basal auditory processing performance. The following three auditory cues were examined, because they are critical for word stress perception: pitch, duration and skewness. The tasks testing for pitch and length discrimination were taken from the Seashore-Test (Stanton, 1928). Skewness discrimination was determined using the procedure developed by (Haake et al., 2013). The procedure was similar to the one used in the Seashore Test. Basically, skewness discrimination required the ability to distinguish the intensity of sounds (stronger vs. weaker). All items were presented via headphones employing Adobe Audition 1.5 (Adobe Systems, San Jose, CA, USA).

**Table 1 | Stimuli for the segmental and suprasegmental conditions.**

	Contrast	Item 1	Item 2
Suprasegmental (SSEG)	PU vs. U	mipá <u>tu</u> tamú <u>pi</u>	mipatú tamupí
	APU vs. U	míkuta kátimu	mikutá katimú
Segmental (SEG)	POA	kúpami mátika	kúpani mátipa
	MOA	kúmita tánuki	kúmifa tánusi

Contrasts are highlighted in bold face. APU, antepenultimate stress; PU, penultimate stress; U, final stress; POA, place of articulation; MOA, combination of place and manner of articulation.



Moreover, given that working memory was crucial for the sequence recall task used in the present study, measures of working memory span were determined for each participant (letter word span forward and backward, following the German version of the Wechsler Memory Scale for number word span forward and backward; Tewes, 1991). Participants were asked to repeat sequences of letters which were given by the examiner. For letter span forward, participants had first to repeat two sequences of three different letters, respectively (for example: f-b-i and c-g-e). At the second level of complexity two sequences of four different letters had to be repeated, respectively, and so forth. On the highest (sixth) level participants had to repeat two sequences of eight letters. For the letter span backward task participants were asked to repeat two sequences of two up to eight letters, respectively, in inverted order. The test procedure was stopped when a participant repeated both sequences on a given level incorrectly.

### fMRI PROCEDURE

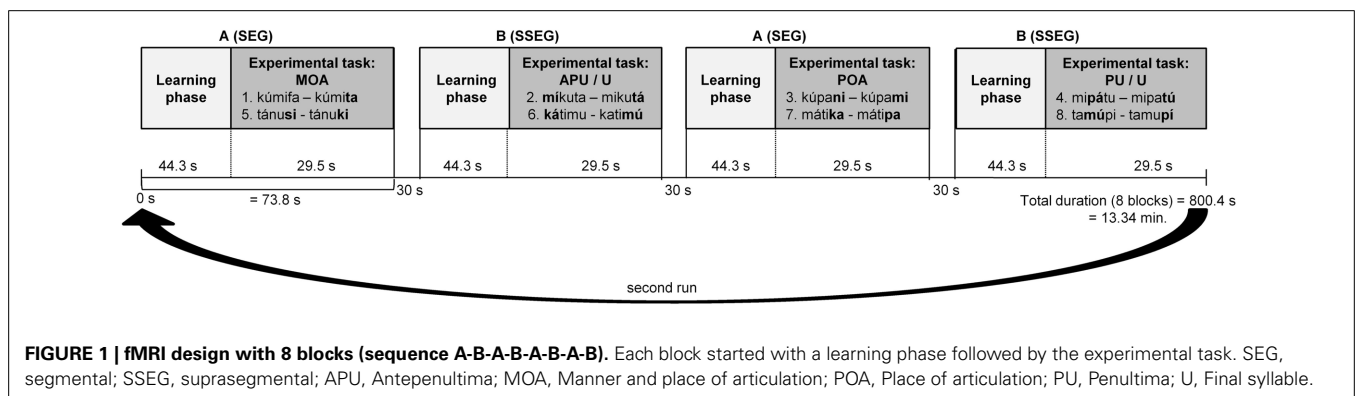
The experiment was a combined behavioral and fMRI study. Participants were lying in the scanner, listening to the pseudowords presented via headphones. They had response boxes in both hands and were instructed to press the correct response buttons with the index finger of the respective hand. Head movements were prevented by using soft foam pads. To familiarize participants with the task and to reduce potential training effects during fMRI data acquisition, all participants were given the opportunity to practice two blocks (one per type of contrast) in a separate room before entering the scanner. The same pseudowords as employed in the scanner served as practice items, but spoken by different speakers (a female native speaker of Dutch and a male native speaker of German).

The experiment had a block design and comprised 8 blocks, each one of which lasted about 73.8 s. Each block consisted of two phases: a learning phase and an experimental phase. There were two types of blocks: Block A contained the segmental condition, and Block B the suprasegmental condition. Blocks were separated by pauses of 30 s. The blocks were presented in an alternating fashion, either starting with Block A (A-B-A-B etc.) or starting with Block B (B-A-B-A etc.), counterbalanced over participants (see Figure 1).

In each learning phase the two pseudowords needed for the following experimental task were presented, such that participants

could familiarize with both words and their association with the respective response button (see Figure 1). Participants were instructed to respond to the first pseudoword encountered by pressing the right button. In this way the right button was always correct for the first pseudoword, such that no further explanation of the correct association between pseudowords and response buttons was needed. When hearing the second pseudoword of the learning phase, participants had to decide whether it matched with the first one (pressing the right button) or not (pressing the left button). Here matching refers to a phonological (type-based) rather than a phonetic (token-based) match. The participants had to make this decision in a sequence of 12 pseudowords per learning phase in pseudorandomized order such that no more than two identical items were presented in a row. The items were spoken either by the male or the female speaker, but no more than three times in a row by the same speaker. Participants were instructed to respond as fast and as accurately as possible by pressing the corresponding button after stimulus presentation. Maximum duration of response time was set to 2000 ms. Only in the learning phase Feedback was presented immediately after each trial only in the learning phase: a “Smiley” for a correct response and a “Frowney” for an incorrect or missing response. The learning phase lasted for about 44.3 s per block. At the end of the learning phase, participants had learned the correct correspondence between both pseudowords and their associated response buttons, which was also valid for the following experimental task.

In the experimental phase participants were presented with pairs of pseudowords from the set of items learned in the preceding learning phase. The task was to press the respective response buttons (as learned in the preceding learning phase) in the order the pseudowords had just been presented. No feedback was provided during the experimental phase. Eight item pairs were presented in random order per block. There were 12 different randomized orders of items for each block, such that only three to four participants had the same order of items. In each item pair, one item was spoken by the male and one by the female speaker. The duration of the experimental phase was 29.5 s per block (see Figure 1). Between pairs in the experimental phase, the background color was slightly modified (a different shade of gray for each sequence) to visually indicate the start of a new pair. Overall, the experiment took 13:34 min. The





experiment was presented with Presentation software (version 14.5, Neurobehavioral Systems, Albany, USA).

### IMAGING ACQUISITION

For each participant, a high-resolution T1-weighted anatomical scan was acquired with a 3T Philips Magnetom MRI system using the standard head coil ( $TR = 9.89$  s, matrix  $256 \times 256$  mm, 176 slices, voxel size =  $1 \times 1 \times 1$  mm<sup>3</sup>;  $FOV = 256$  mm,  $TE = 4.59$  ms; flip angle =  $8^\circ$ ). Moreover, one functional imaging block sensitive to blood oxygenation level-dependent (BOLD) contrast was recorded for each participant ( $T2^*$ -weighted echo-planar sequence,  $TR = 2.89$  s;  $TE = 30$  ms; flip angle =  $79^\circ$ ;  $FOV = 240$  mm;  $80 \times 80$  matrix; 42 slices, voxel size =  $3 \times 3 \times 3$  mm<sup>3</sup>, gap = 0.5 mm).

### ANALYSIS OF BEHAVIORAL DATA

Behavioral data analysis was based responses in the experimental phase only. Furthermore, items with response latency faster than 200 ms were not considered. Analyses focused on accuracy data since reaction times in the suprasegmental condition were confounded with different “points of uniqueness” when participants were able to detect the stress difference in a pair of pseudowords (e.g., earlier point of uniqueness in “míkuta” vs. “míkúta” compared to “míkuta” vs. “mikutá”).

Participant's individual performance in word stress processing was evaluated employing accuracy data of the suprasegmental condition. Based on a median split of the number of correct trials in the suprasegmental condition (see **Figure 2**), each participant was assigned either to a group of poor performers (below average) or to a group of good performers (above average).

In an initial step, a  $2 \times 2$  repeated measures Analysis of Variance (ANOVA) on accuracy was performed with the within-participant factor condition (segmental vs. suprasegmental) and

the between-participant factor group (above vs. below average word stress processing).

To pursue the potential association between performance in basal auditory processing, working memory, and suprasegmental processing, a stepwise multiple regression analysis with mean accuracy in the suprasegmental condition as criterion variable was conducted, which was stopped when the inclusion of another predictor would not increase  $R^2$  significantly (at  $p < 0.05$ ). The predictors incorporated were performance measures from the pretest tasks, i.e., pitch discrimination, duration discrimination, skewness discrimination, a combined measure of these three auditory processing tasks (mean auditory processing accuracy), and working memory span.

### ANALYSIS OF IMAGING DATA

The anatomical scans were normalized and averaged in SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>). The fMRI time series were corrected for movement in SPM8. Images were motion corrected and realigned to each participant's first image. Data was normalized into standard MNI space. Images were resampled every 2.5 mm using 4th degree spline interpolation and smoothed with a 6 mm FWHM Gaussian kernel to accommodate inter-subject variation in brain anatomy and to increase signal-to-noise ratio in the images. The data were high-pass filtered (128 s) to remove low-frequency signal drifts and corrected for autocorrelation assuming an AR(1) process. Brain activity was convolved over all experimental trials with the canonical haemodynamic response function (HRF) and its derivative.

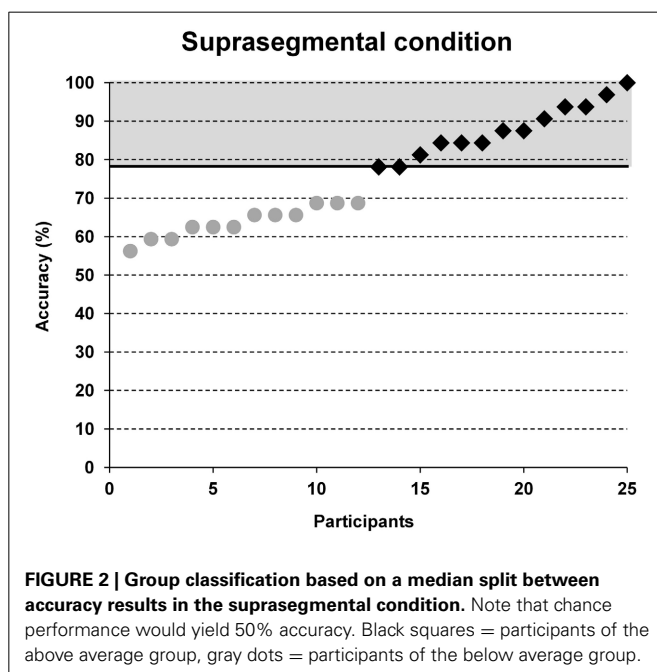
On the first level, the intraindividual beta contrast weights for segmental and suprasegmental processing were evaluated. On the second level, both main effects and their interaction were evaluated in a  $2 \times 2$  (flexible factorial) ANOVA with the between-subject factor group (above vs. below average) and the within-participant factor condition (segmental vs. suprasegmental). For the anatomical localization of effects, the anatomical automatic labeling tool (AAL) in SPM8 (<http://www.cyceron.fr/index.php/en/plateforme-en/freeware>) was used to identify Brodmann Areas (BA). If possible, the SPM Anatomy Toolbox (Eickhoff et al., 2005), available for all published cytoarchitectonic maps from [www.fz-juelich.de/ime/spm\\_anatomy\\_toolbox](http://www.fz-juelich.de/ime/spm_anatomy_toolbox), was additionally used and in the results will be indicated by an “Area” specification.

## RESULTS

### BEHAVIORAL DATA

Accuracy in the segmental task ranged from 56.3 to 96.9% and in the suprasegmental task from 56.3 to 100%. The group classification was based on a median split for the accuracy results in the suprasegmental condition (see **Figure 2**). The ratio of male and female participants was comparable between both groups (good: 8m/5f, poor: 8m/4f). A descriptive overview of the results is provided in **Figure 3**.

A repeated measures ANOVA over arcsine-transformed error rates revealed a significant main effect of group [ $F_{(1, 23)} = 12.16$ ,  $p < 0.01$ ], indicating that good performers made significantly less errors (in total) than poor performers (16.0 vs. 29.0%, see **Figure 3**). There was no main effect of condition [ $F_{(1, 23)} < 1$ ].



However, there was a significant two-way interaction of condition and group [ $F_{(1, 23)} = 9.3, p < 0.01$ ]. The effect of condition was only significant for poor performers [ $t_{(11)} = 3.24; p < 0.01$ ], meaning that in this group the error rate in the suprasegmental condition was higher than in the segmental condition (36.2 vs. 21.9%). In contrast, for good performers the effect of condition did not reach significance [ $t_{(12)} = 1.78, p = 0.10$ ]. However, it should be noted that, in contrast to the poor performers, error rate was numerically higher in the segmental than in the suprasegmental condition (19.7 vs. 12.3%).

Crucially, both groups differed significantly only in the suprasegmental condition [ $t_{(24)} = 6.21, p < 0.001$ ] indicating that the good performers performed reliably better (87.7%) than the poor performers (63.8%). There was no significant difference between groups for the segmental condition [ $t_{(24)} = -0.4, p = 0.70$ ], see **Figure 3**. Furthermore, no correlation was observed between stress processing (suprasegmental) and consonant processing (segmental) (Spearman rho = 0.072,  $p = 0.733$ ).

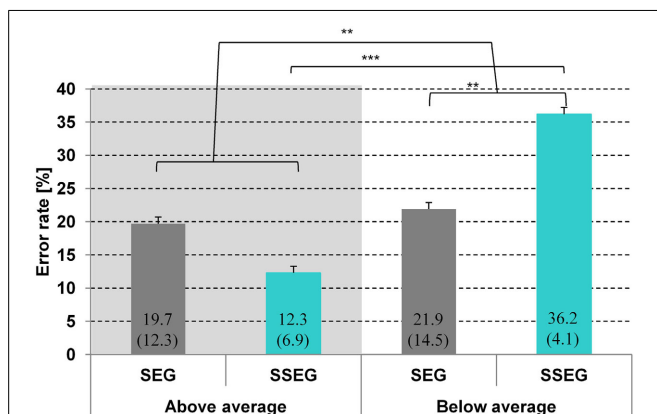
In order to examine whether performance in the suprasegmental condition was influenced by basic auditory processing abilities and/or working memory skills, a stepwise multiple linear regression analysis was performed over arcsine-transformed error rate of the suprasegmental condition. The final model comprised the predictors auditory processing and working memory span forward [ $R^2 = 0.400$ , adjusted  $R^2 = 0.345$ ,  $F_{(2, 24)} = 7.3$ ,  $p < 0.01$ ].

## fMRI DATA

Analysis of fMRI data was based on all trials in the experimental phase. In a first step, a conjunction analysis was conducted to identify common overall activation in the paradigm irrespective of group and condition.

## OVERVIEW: CONJUNCTION ANALYSIS

A conjunction over all conditions and groups was calculated (SEG in poor performers, SSEG in poor performers, SEG in good performers, SSEG in good performers) to show joined activation at an uncorrected voxelwise  $p < 0.0001$ . Please note that



**FIGURE 3 | Comparison of mean error rate (%) per condition and group.** Standard deviations are given in parentheses. SEG, segmental condition; SSEG, suprasegmental condition. \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ .

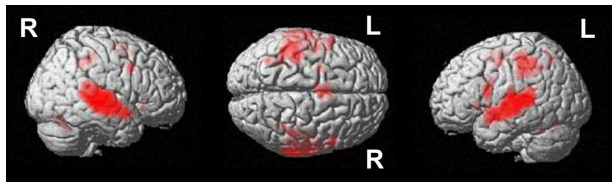
this more rigorous  $p$ -value had to be used in the conjunction (compared to the level of  $p < 0.001$  for the complex contrasts reported below) to visualize the different maxima of activation (cf. Wood et al., 2009; Klein et al., 2010). However, all activations reported here remain significant following family-wise error correction (FWE) at a cluster-level of  $p < 0.05$ . Significant activations in the entire primary auditory cortex were present (see **Table 2** and **Figure 4**). Bilateral activation was found in the superior temporal gyrus/sulcus (STG; STS) and the middle temporal gyrus (MTG). Furthermore, left-hemispheric clusters of activated voxels were observed in the inferior frontal gyrus (IFG; Area 44, Area 6 (BA 44); SPM Anatomy Toolbox, Amunts et al., 1999; cf. Eickhoff et al., 2005), the insula, the inferior parietal sulcus (IPS; hIP2, IPC (PF, PFm), hIP1 (BA 7); SPM Anatomy Toolbox, Choi et al., 2006; cf. Eickhoff et al., 2005) the SMA, and the middle frontal gyrus (MFG). In the right hemisphere voxels in the IFG, inferior parietal lobule (hIP2, SPL (7PC), hIP1, hIP3; SPM Anatomy Toolbox, Choi et al., 2006; Scheperjans et al., 2008a,b; cf. Eickhoff et al., 2005) and the cerebellum were activated, while the precentral gyrus was found active bilaterally (see **Table 2**, **Figure 4**).

**Table 2 | Maxima of the conjunction analysis over both conditions (segmental and suprasegmental) as well as both groups (above and below average) at an uncorrected voxelwise  $p < 0.0001$  (cluster-corrected FWE of  $p < 0.05$ ).**

Brain region (BA)	MNI			Cluster size	z score
	x	y	z		
RH superior temporal gyrus/sulcus (BA 22)	57	-19	-2	963	7.62
RH middle temporal gyrus (BA 21) <sup>a</sup>	57	-28	-5		7.06
LH superior temporal gyrus (BA 41/42) <sup>a</sup>	-51	-22	4		6.52
LH middle temporal gyrus (BA 22)	-60	-31	4	885	7.13
LH insula	-33	20	1	13	4.71
LH inferior frontal gyrus (BA 44)	-57	8	16	55	5.40
RH inferior frontal gyrus (BA 47)	33	26	-5	11	4.54
LH SMA (BA 6)	-3	-4	58	82	4.57
LH IPS (BA 7)	-45	-43	43	293	5.30
RH IPS (BA 7)	39	-43	43	41	4.84
LH precentral gyrus (BA 6)	-48	-4	46	11	4.56
RH precentral gyrus (BA 6)	57	11	37	17	4.49
LH middle frontal gyrus (BA 6)	-27	-4	52	17	4.37
RH cerebellum	6	-67	-20	58	4.82

IPS, inferior parietal sulcus; LH, left hemisphere; RH, right hemisphere; SMA, supplementary motor area.

<sup>a</sup>Minor maximum.



**FIGURE 4 |** Conjunction analysis over all groups and conditions at an uncorrected voxelwise  $p < 0.0001$  (cluster-corrected FWE of  $p < 0.05$ ).

## CONDITION-BASED COMPARISONS

### Suprasegmental vs. segmental processing

Suprasegmental was contrasted to segmental processing at an uncorrected voxelwise threshold of  $p < 0.001$  and a cluster size of  $k = 10$  voxels (see **Figure 5A**, **Table 3**). Larger activation for suprasegmental processing was found bilaterally in the IFG (Area 44 and Area 45 (BA44 and BA 45); SPM Anatomy Toolbox, cf. Eickhoff et al., 2005) as well as in the insula. Furthermore, in the left hemisphere the thalamus, the IPS (hIP1, hIP3 (BA 7); SPM Anatomy Toolbox, cf. Eickhoff et al., 2005) and the pre-SMA (BA 6) were activated, while in the right hemisphere the pallidum as well as the right SMA (BA 6) revealed stronger activation in stress processing compared to consonant processing. Further clusters of activated voxels were found in the bilateral precentral gyrus, in the left MFG (BA 10) and in the cerebellum, bilaterally.

### Segmental vs. suprasegmental processing

Inspection of the inverse contrast (uncorrected  $p < 0.001$ ,  $k = 10$  voxel) revealed activation in the bilateral SMA (BA 6), the right middle orbital gyrus and the left precuneus (see **Figure 5B**, **Table 3**).

## GROUP-BASED COMPARISONS

### Poor performers vs. good performers

Poor performers revealed significantly stronger activation than good performers in the left MTG at an uncorrected voxelwise  $p < 0.001$  and a cluster size of 10 voxels (see **Figure 6A**, **Table 3**).

### Good performers vs. poor performers

When comparing good performers vs. poor performers (uncorrected  $p < 0.001$ ,  $k = 10$  voxel), significantly more activation was found in the left precuneus (see **Figure 6B**, **Table 3**).

## INTERACTION BETWEEN GROUP AND CONDITION

We conducted an ANCOVA over participants on the fMRI data with working memory and auditory performance from the pretest as covariates, to correct the segmental and suprasegmental activations for working memory and auditory abilities. In this context, we also examined whether there is additional fMRI variance, which is exclusively explained by the covariates. However, at the threshold given (FWE-cluster threshold corrected) there was no such additional activity to be found.

Group and condition interacted significantly in the right hippocampus (CA (BA 27), SPM Anatomy Toolbox, Amunts et al., 2005; cf. Eickhoff et al., 2005) and cerebellum at an uncorrected voxelwise  $p < 0.001$  and a cluster size of 10 voxels (see

**Figure 7**, **Table 3**). However, especially in the cerebellum the interactions in signal change seem to be mostly due to different degrees of deactivation. However, it can be seen that good performers showed relatively more activation (or less deactivation, respectively) in the segmental condition in the right hippocampus and cerebellum compared to poor performers, whereas poor performers revealed relatively stronger activation compared to good performers in these areas in the suprasegmental condition.

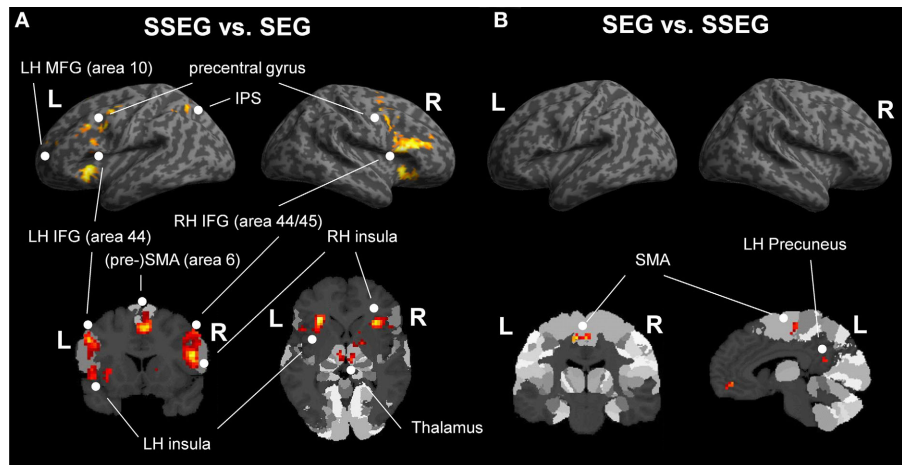
## DISCUSSION

The current study set off to examine whether there are interindividual differences in word stress processing performance in native speakers of German and, if so, which neural correlates underlie these differences. So far, most studies focused on typologically motivated processing differences between speakers of languages with fixed vs. variable stress. In particular, Dupoux, Peperkamp and colleagues compared speakers of Spanish (variable stress pattern) to speakers of French (fixed stress pattern; see Dupoux et al., 1997, 2001; Peperkamp et al., 1999; Peperkamp and Dupoux, 2002) and found superior performance of the former compared to the latter (for similar results in a comparison between French and German see Schmidt-Kassow et al., 2011a). Interindividual differences within one language—although repeatedly observed—were treated as noise (Peperkamp et al., 1999; Domahs et al., 2008, 2013b; Dupoux et al., 2010) or were left unexplained (Boecker et al., 1999).

In the present study, participants were examined in both suprasegmental as well as segmental variants of the sequence recall task both at a behavioral and at a neuro-functional level. Indeed, based on behavioral results we were able to identify considerable interindividual differences within native speakers of German (accuracy in the suprasegmental task ranging from floor to ceiling performance).

To explore more thoroughly, which factors modulate suprasegmental processing differences, working memory span as well as auditory processing abilities were analyzed. In fact, we demonstrated that suprasegmental performance was predicted by both basic auditory processing abilities (i.e., duration, time, skewness discrimination) and working memory span. The influence of working memory on performance in the suprasegmental task seems highly plausible since working memory was clearly task-relevant. Crucially, the fact that a combined measure of duration, time, and skewness discrimination predicted individual performance in word stress processing, provides a first hint toward an explanation for the interindividual variability observed. This result fits nicely with findings recently reported by Haake et al. (2013), who observed that word stress processing in children with specific language impairment as well as in typically developing children is predicted by auditory processing of duration cues. Obviously, basic auditory processing performance may exert its influence not only in children, but also in healthy adults for whom the recognition and interpretation of word stress is relevant in their native language.

In sum, there was substantial interindividual variability in word stress processing. Hence, two groups were defined based



**FIGURE 5 | (A)** Comparison of suprasegmental vs. segmental condition (uncorrected  $p < 0.001$ ,  $k = 10$  voxels). **(B)** Segmental vs. suprasegmental condition (uncorrected  $p < 0.001$ ,  $k = 10$  voxels). IFG, inferior frontal gyrus;

IPS, inferior parietal sulcus; LH, left hemisphere; MFG, middle frontal gyrus; RH, right hemisphere; SEG, segmental; SMA, supplementary motor area; SSEG, suprasegmental.

**Table 3 | Significant brain activation differences for various group and condition contrasts.**

Contrast	Brain region (BA)	MNI			Cluster size	z score
		x	y	z		
Suprasegmental vs. segmental	RH inferior frontal gyrus (BA 44/45)	51	17	16	345	5.19
	RH inferior frontal gyrus (BA 45) <sup>a</sup>	54	29	22		4.97
	LH inferior frontal gyrus (BA 44/45)	-51	14	31	112	4.38
	LH insula	-30	20	-11	115	5.15
	RH insula	33	23	-2	134	4.50
	LH thalamus	-6	-13	-2	151	3.85
	RH pallidum	18	-1	1	70	4.02
	LH IPS [hIP1, hIP3 (BA 7)]	-36	-52	37	79	4.08
	LH pre-SMA (BA 6)	0	17	46	184	4.95
	RH SMA (BA 6)	9	2	58	12	3.67
	RH precentral gyrus (BA 6)	42	-1	49	100	4.37
	LH precentral gyrus (BA 6)	-39	-4	40	55	4.31
	LH middle frontal gyrus (BA 10)	-30	53	22	16	3.94
	LH cerebellum	-9	-76	-29	140	4.05
	RH cerebellum	24	-34	-41	31	4.01
Segmental vs. suprasegmental	LH SMA (BA 6)	-9	-19	52	44	3.98
	RH SMA (BA 6)	15	-10	40	19	4.09
	RH middle orbital gyrus (BA 10)	3	50	-5	88	4.19
	LH precuneus (BA 7)	-6	-58	19	13	3.47
Below vs. above	LH middle temporal gyrus (BA 21)	-69	-34	-8	16	3.84
Above vs. below	LH precunes (BA 7)	0	-52	40	24	3.71
Interaction group * condition	RH hippocampus	30	-34	-2	11	3.76
	RH cerebellum	24	-31	-23	16	4.48

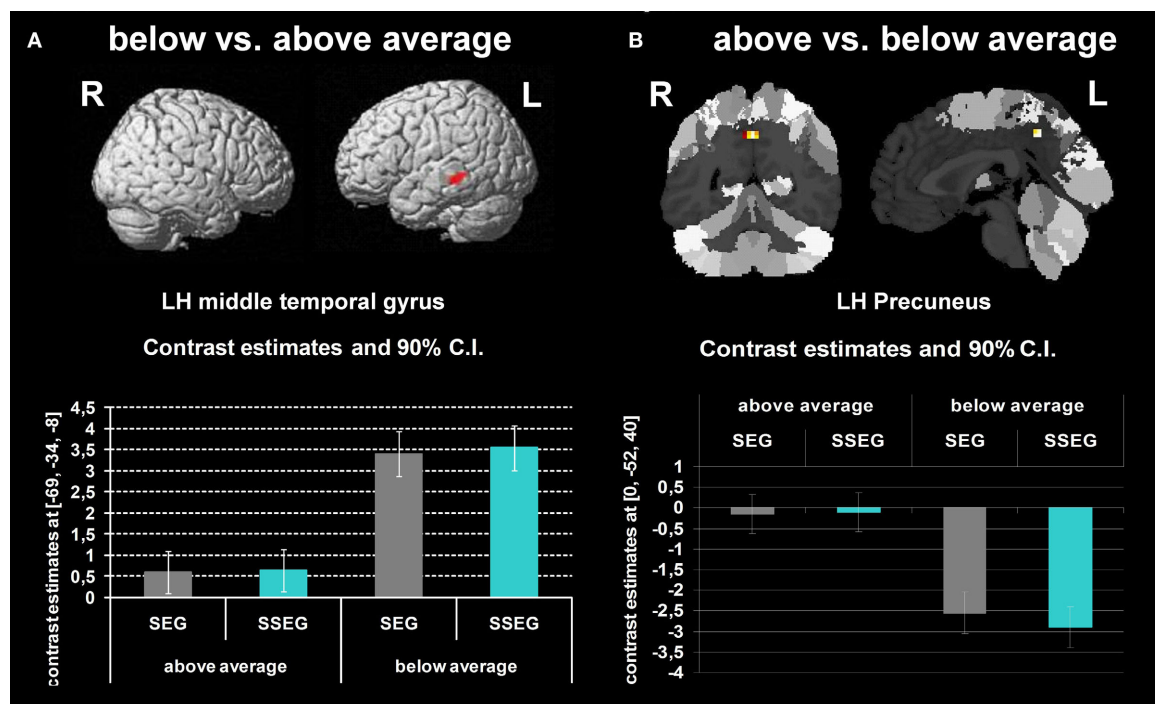
IPS, inferior parietal sulcus; LH, left hemisphere; RH, right hemisphere; SMA, supplementary motor area.

<sup>a</sup>Minor maximum.

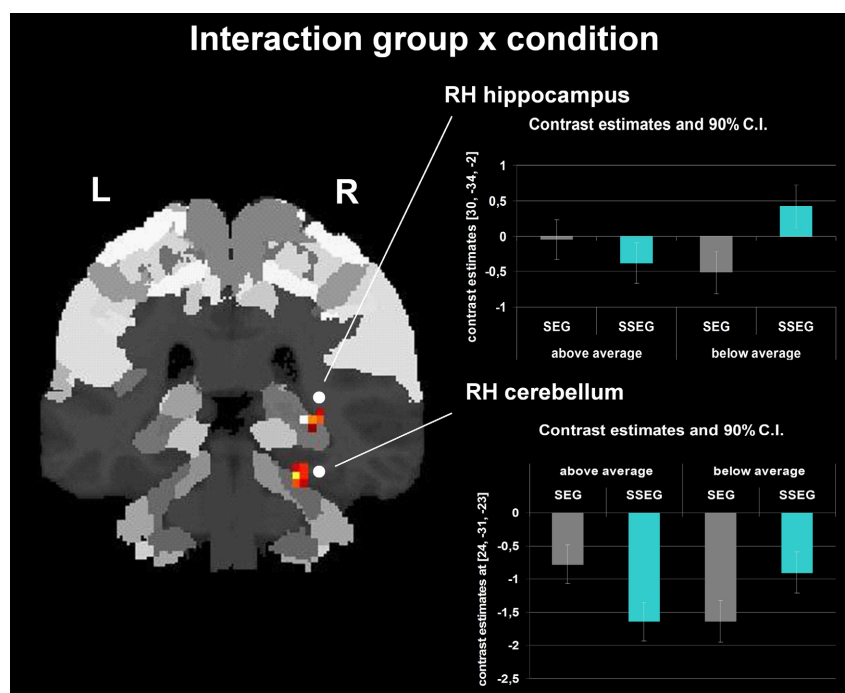
on a median split of individual accuracy results in the suprasegmental task. Neural correlates of segmental and suprasegmental processing and their interaction with group membership were investigated and will be discussed in the following.

## NEURAL CORRELATES OF SEGMENTAL AND SUPRASEGMENTAL PROCESSING

The conjunction analysis revealed a large cluster of activation in auditory cortex across performance levels and conditions



**FIGURE 6 | (A)** Comparison of participants below vs. above average (uncorrected  $p < 0.001$ ,  $k = 10$  voxels). **(B)** Participants above vs. below average (uncorrected  $p < 0.001$ ,  $k = 10$  voxels). The bar charts below the activation figure depict the corresponding beta estimates for the respective brain region.



**FIGURE 7 | Interaction between group and condition (uncorrected  $p < 0.001$ ,  $k = 10$  voxels).** The bar charts next to the activation figure depict the corresponding beta estimates for the respective brain region.



(cf. **Figure 4**, **Table 2**), extending from the superior temporal gyrus to the middle temporal gyrus and to the insula. This finding is highly plausible, because participants had to process auditory linguistic stimuli. More specifically, previous studies reported activation in the STG or STS for processing of prosodic information in general (e.g., Dogil, 2003; Ischebeck et al., 2008), and for processing of word stress in particular (Aleman et al., 2005; Klein et al., 2011; Domahs et al., 2013b).

In addition, activation in the bilateral supplementary motor area (with left-hemispheric peak activation within a large cluster extending into the right hemisphere) and in the bilateral inferior parietal sulcus was found. This may be related to the fact that participants had to determine either stress localization or consonant differences by button presses since the SMA has been suggested to subserve decision making (Kong et al., 2005). Additionally, a combination of working memory related BA 44 and intraparietal BA 7 activation indicated that participants had to hold the sequences of pseudowords in working memory. Moreover, bilateral activation in the precentral gyrus was observed, probably indicating motor processing associated with finger movements and button presses (Zilles and Rehkaemper, 1998).

Beyond these task-related effects, cerebellum, temporal cortex, premotor cortex, preSMA/SMA and inferior frontal cortex have been described as part of a network involved in speech perception, especially engaged in the temporal processing of speech (Grahn and Brett, 2007; Kotz et al., 2009; Kotz and Schwartz, 2010).

### SUPRASEGMENTAL vs. SEGMENTAL PROCESSING

In the behavioral data, no correlation was observed between stress processing (suprasegmental) and consonant processing (segmental). This suggests that the linguistic abilities underlying these two conditions may be to a certain degree independent, although they were tested with a comparable paradigm in the present study.

When the suprasegmental task was contrasted to the segmental task, a subcortico-cortico-cerebellar network of brain regions was revealed, including bilateral IFG (BA44 and BA 45), bilateral insula, bilateral precentral gyrus, bilateral cerebellum, left thalamus, left pre-SMA (BA 6), right globus pallidus, and right SMA (BA 6). There is accumulating evidence, that this network is involved in processing spectro-temporal aspects of speech (Lutz et al., 2000; Lewis et al., 2004; Bengtsson et al., 2005; Riecker et al., 2006; Grahn and Brett, 2007; Coull et al., 2008; Geiser et al., 2008; Kotz et al., 2009; Kotz and Schwartz, 2011; Schwartz et al., 2012a,b, see Kotz and Schwartz, 2010, for a review). This finding seems very plausible, given that duration is the most relevant acoustic cue to word stress in German (Jessen and Marasek, 1997; Classen et al., 1998; Schneider, 2007; Schneider and Möbius, 2007; Lintfert, 2010) and performance in auditory discrimination in general and duration discrimination in particular predicts performance in the more complex task related to word stress (behavioral results of the present study, see Haake et al., 2013, for evidence from German speaking children).

More specifically, bilateral activation in the inferior frontal gyri related to the suprasegmental condition is in line with previous

studies, which reported these areas to be activated in processing linguistic aspects of prosody (e.g., Wildgruber et al., 2004; Li et al., 2010; Klein et al., 2011; Domahs et al., 2013b).

Furthermore, activation in the left insula related to suprasegmental processing is consistent with previous studies, which found this area activated for auditory temporal processing (Lewis et al., 2000; Ackermann et al., 2001; Lewis and Miall, 2003), for pitch-related stimuli (Zarate and Zatorre, 2005) as well as for auditory timing perception (Geiser et al., 2008) and word stress processing proper (Aleman et al., 2005; Klein et al., 2011).

Activation in the bilateral inferior parietal sulcus may reflect the fact that participants had to store information in working memory and to respond by button presses. Possibly, they employed a spatial representation of the pseudowords (e.g., first syllable = left, last syllable = right) and of response buttons to come to the correct decision. Amongst others, the intraparietal cortex has been suggested to subserve mental imagery (Just et al., 2004). Moreover, the IPS has been frequently reported to be involved in the processing of proximity relations (see Dehaene et al., 2003 for a review). Recall that stress is an inherently relational property and requires the comparison of acoustic cues (e.g., duration, pitch, and skewness) between stressed and unstressed syllables. In the present study, the inferior parietal sulcus may be associated with mental imagery and with the evaluation of gradual differences in acoustic cues related to word stress. This might comprise positional information, which has to be encoded in the IPS and held in working memory as well as the actual comparison process of the positional information within the sequences of CV-syllables—a process also most probably associated with the intraparietal cortices (cf. Klein et al., 2011). In particular, bilateral inferior parietal cortex has been found activated in tasks tapping on suprasegmental compared to segmental aspects of words (Li et al., 2010; Klein et al., 2011).

Beyond temporal processing of speech input, activation in the supplementary motor area may be related to the fact that in general the suprasegmental task in this study was somewhat more difficult than the segmental task. The SMA has been found to support operation procedures (Kong et al., 2005). Interestingly, Domahs et al. (2013b) observed increased activation in bilateral SMA in a difficult compared to an easy condition in a word stress violation task. Moreover, SMA activation in the suprasegmental condition together with a significantly increased activation in the precentral gyrus could point to an involvement of the central motor system. Given that both the SMA and the precentral gyrus were activated bilaterally, these findings may reflect control of finger movements in participants (e.g., Shibasaki et al., 1993; Catalan et al., 1998). Possibly, participants may have needed higher control of their finger movements in the more difficult suprasegmental condition. An alternative explanation could be that in more difficult conditions participants may establish a correspondence between their fingers and the positional information of stress, for instance, by using finger counting. This would be also in line with the activation pattern observed in SMA, precentral and intraparietal areas. However, this account remains speculative so far and needs further evaluation in future studies.



## INTERINDIVIDUAL DIFFERENCES

The middle temporal gyrus was found activated in both conditions (segmental, suprasegmental) for both groups (cf. **Table 3**). This fits well with the fact that the MTG has been associated with phonology (Graves et al., 2010) and, more generally, with complex sound and speech processing (Scott et al., 2000). Nevertheless, poor performers showed stronger activation in this region.

Further significant changes in the BOLD signal were found in the precuneus. These findings are rather difficult to interpret since for good performers the BOLD signal in the precuneus seemed to be close to zero in both the segmental and the suprasegmental conditions (see **Figure 6B**), whereas in poor performers the precuneus was strongly deactivated in both conditions. Considering that the amplitude of the BOLD signal indicated by SPM is subject to arbitrary factors (such as the definition of the baseline), the present findings can only be interpreted in relative terms, not in terms of “activation” or “deactivation.” Generally, the precuneus has been suggested not only to subserve learning of motor-sequences (Sadato et al., 1996; Sakai et al., 1998) but also to be involved in mental imagery (Dehaene et al., 1996; Huijbers et al., 2011). Possibly, good performers may have relied more on mental imagery or motor-sequence learning to solve the task correctly, compared to poor performers. Nevertheless, we are well aware of the fact that currently this explanation remains speculative.

One may conclude that both groups activated the MTG for phonological processing of stimuli in both conditions, but that poor performers required more resources. It may be speculated that good performers have used a combination of visual and auditory representations to solve the tasks, whereas poor performers only relied on auditory information (but to a higher degree). Possibly, a combination of visual and auditory processing may be advantageous.

Although native speakers of German are highly familiar with the use of suprasegmental features in their mother tongue, the present study shows that their performance in an experimental task tapping on this aspect of language may nevertheless be very heterogeneous. Until now, it was assumed that native German speakers should be “naturally” competent in word stress processing, since this is a relevant feature of their language, which is acquired early. Preverbal infants learn the typical stress pattern of their mother tongue and can use it in speech segmentation (Hoehle et al., 2009). Importantly, even those participants, who showed poor performance in the specific suprasegmental task in the present study, were competent speakers of German. Note that the stress pattern of real words is stored in the lexicon. However, in the present study, participants had to process pseudowords which by definition cannot be stored in the mental lexicon. Thus, processing word stress in everyday language requires lexical retrieval, whereas the suprasegmental task in our experiment may have required other types of prosodic knowledge (e.g., rule-based knowledge). Furthermore, every-day language is typically embedded in a redundant context, which helps in resolving ambiguities related to word stress, e.g., in the interpretation of minimal pairs. Therefore, the specific difficulties in suprasegmental processing of pseudowords observed in

the present study are subclinical with no obvious impact on language use.

## INTERACTION BETWEEN GROUP AND CONDITION

Behaviorally, a two-way interaction of condition (segmental vs. suprasegmental) and group (below vs. above average) indicated that the good performers were numerically better in suprasegmental than in segmental processing, whereas the poor performers were significantly better in segmental than in suprasegmental processing (see **Figure 3**). Importantly, a two-way interaction of condition and group was also revealed in the neuro-functional data (see **Figure 7, Table 3**). Good performers showed relatively more activation (or less deactivation, respectively) in the segmental condition in the right hippocampus and cerebellum compared to poor performers, whereas poor performers revealed relatively stronger activation in these areas in the suprasegmental condition compared to good performers.

Hippocampal cells have been shown to be involved in auditory working memory in rats (Sakurai, 1990, 1994). More recently, the hippocampus has been argued to contribute to performance in a variety of cognitive tasks including working memory and perception, when these tasks require high-resolution binding of features and relational information (Yonelinas, 2013). Clearly, the sequence recall task used in the present experiment does require such a complex and demanding type of binding. Interestingly, activation in the right hippocampus was related to relative task difficulty: Poor performers seemed to need relatively more cognitive resources in the suprasegmental task (which they performed worse than the segmental task), but good performers seemed to put relatively more effort into the segmental task (which they performed worse than the suprasegmental task).

Furthermore, a similar pattern of (de-)activation was observed for the interaction in the right cerebellum. The cerebellum has been considered to be part of a network related to the processing of spectro-temporal aspects of speech (Kotz and Schwartz, 2010). The interaction in the cerebellum suggests that poor performers may have needed the cerebellum relatively more for the suprasegmental task (although achieving inferior results) than good performers. The opposite pattern was observed in the segmental condition. Again, these interpretations have to be considered very cautiously and remain speculative, because the interaction pattern consists only of different degrees of deactivation.

## CONCLUSION AND PERSPECTIVES

The present study is a first step toward a more comprehensive understanding of the processing of word stress. In particular, it highlights the need to examine brain activation data not only at the second level in group analyses, but also to analyze individual data at the first level. Taken together, our results provide behavioral and neuro-functional evidence for substantial interindividual differences within a group of native speakers of German, a language with variable stress, in word stress processing. They suggest that part of the behavioral variance is explained by basic auditory processing and working memory performance. It would be interesting to explore, whether speakers of a language with fixed stress (e.g., Czech, Finnish, Polish, Turkish, Persian, or French) show similar interindividual heterogeneity.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00365/abstract>

## REFERENCES

- Ackermann, H., Riecker, A., Mathiak, K., Erb, M., Grodd, W., and Wildgruber, D. (2001). Rate-dependent activation of a prefrontal–insular–cerebellar network during passive listening to trains of click stimuli: an fMRI study. *Neuroreport* 12, 4087–4092. doi: 10.1097/00001756-200112210-00045
- Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Haan, E. H. F., and Kahn, K. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cereb. Cortex* 15, 221–228. doi: 10.1093/cercor/bhh124
- Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N. J., et al. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat. Embryol.* 210, 343–352. doi: 10.1007/s00429-005-0025-5
- Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H. B., and Zilles, K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.* 412, 319–341.
- Bengtsson, S. L., Ehrsson, H. H., Forssberg, H., and Ullen, F. (2005). Effector-independent voluntary timing: behavioral and neuroimaging evidence. *Eur. J. Neurosci.* 22, 3255–3265. doi: 10.1111/j.1460-9568.2005.04517.x
- Boecker, K. B. E., Bastiaansen, M. C. M., Vroomen, J., Brunia, C. H. M., and De Gelder, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology* 36, 706–720. doi: 10.1111/1469-8986.3660706
- Catalan, M. J., Honda, M., Weeks, R. A., Cohen, L. G., and Hallett, M. (1998). The functional neuroanatomy of simple and complex sequential finger movements: a PET study. *Brain* 121(Pt 2), 253–264. doi: 10.1093/brain/121.2.253
- Choi, H. J., Zilles, K., Mohlberg, H., Schleicher, A., Fink, G. R., Armstrong, E., et al. (2006). Cytoarchitectonic identification and probabilistic mapping of two distinct areas within the anterior ventral bank of the human intraparietal sulcus. *J. Comp. Neurol.* 495, 53–69. doi: 10.1002/cne.20849
- Classen, K., Dogil, G., Jessen, M., Marasek, K., and Wokurek, W. (1998). Stimmqualität und Wortbetonung im Deutschen. *Linguist. Berichte* 174, 202–245.
- Coull, J. T., Nazarian, B., and Vidal, F. (2008). Timing, storage, and comparison of stimulus duration engage discrete anatomical components of a perceptual timing network. *J. Cogn. Neurosci.* 20, 2185–2197. doi: 10.1162/jocn.2008.20153
- Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239
- Dehaene, S., Tzourio, N., Frak, V., Raynaud, L., Cohen, L., Mehler, J., et al. (1996). Cerebral activations during number multiplication and comparison: a PET study. *Neuropsychologia* 34, 1097–1106. doi: 10.1016/0028-3932(96)00027-9
- Dogil, G. (2003). “Understanding prosody,” in *Psycholinguistik: Ein Internationales Handbuch*, eds H. Rickheit, T. Herrmann, and W. Deutsch (Berlin: de Gruyter), 544–565.
- Domahs, U., Genc, S., Knaus, J., Wiese, R., and Kabak, B. (2013a). Processing (un-)predictable word stress: ERP evidence from Turkish. *Lang. Cogn. Process.* 28, 335–354. doi: 10.1080/01690965.2011.634590
- Domahs, U., Klein, E., Huber, W., and Domahs, F. (2013b). Good, bad and ugly word stress – fMRI evidence for foot structure driven processing of prosodic violations. *Brain Lang.* 125, 272–282. doi: 10.1016/j.bandl.2013.02.012
- Domahs, U., Knaus, J., Orzechowska, P., and Wiese, R. (2012). Stress “deafness” in a language with fixed word stress: an ERP study on Polish. *Front. Lang. Sci.* 3:439. doi: 10.3389/fpsyg.2012.00439
- Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I., and Schlewsky, M. (2008). The processing of German word stress: evidence for the prosodic hierarchy. *Phonology* 25, 1–36. doi: 10.1017/S0952675708001383
- Dupoux, E., Pallier, C., Sebastian, N., and Mehler, J. (1997). A destressing “Deafness” in French? *J. Mem. Lang.* 36, 406–421. doi: 10.1006/jmla.1996.2500
- Dupoux, E., Peperkamp, S., and Sebastián-Gallés, N. (2001). A robust method to study stress “deafness.” *J. Acoust. Soc. Am.* 110, 1606–1618. doi: 10.1121/1.1380437
- Dupoux, E., Peperkamp, S., and Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: stress ‘deafness’ in simultaneous French-Spanish bilinguals. *Cognition* 114, 266–275. doi: 10.1016/j.cognition.2009.10.001
- Dupoux, E., Sebastián-Gallés, N., Navarrete, E., and Peperkamp, S. (2008). Persistent stress ‘deafness’: the case of French learners of Spanish. *Cognition* 106, 682–706. doi: 10.1016/j.cognition.2007.04.001
- Eickhoff, S., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034
- Eisenberg, P. (1991). Syllabische Struktur und Wortakzent: Prinzipien der Prosodik deutscher Wörter. *Zeitschrift für Sprachwissenschaft*, 10, 37–64. doi: 10.1515/zfsw.1991.10.1.37
- Eisenberg, P. (2006). *Grundriss der deutschen Grammatik: Das Wort*, 3rd Edn. Stuttgart: Metzler.
- Féry, C. (1998). German word stress in optimality theory. *J. Comp. Ger. Linguist.* 2, 101–142. doi: 10.1023/A:1009883701003
- Geiser, E., Zaehle, T., Jancke, L., and Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by meter processing: an fMRI study. *J. Cogn. Neurosci.* 20, 541–552. doi: 10.1162/jocn.2008.20029
- Giegerich, H. J. (1985). *Metrical Phonology and Phonological Structure: German and English*. Cambridge: Cambridge University Press.
- Grahn, J. A., and Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* 19, 893–906. doi: 10.1162/jocn.2007.19.5.893
- Graves, W. W., Desai, R., Humphries, C., Seidenberg, M. S., and Binder, J. R. (2010). Neural systems for reading aloud: a multiparametric approach. *Cereb. Cortex* 20, 1799–1815. doi: 10.1093/cercor/bhp245
- Haake, C., Kob, M., Willmes, K., and Domahs, F. (2013). Word stress processing in specific language impairment: auditory or representational deficits? *Clin. Linguist. Phon.* 27, 594–615. doi: 10.3109/02699206.2013.798034
- Heim, S., and Alter, K. (2006). Prosodic pitch accents in language comprehension and production: ERP data and acoustic analyses. *Acta Neurobiol. Exp.* 66, 55–68.
- Heim, S., and Alter, K. (2007). “Focus on focus: the brain's electrophysiological response to focus particles and accents in German,” in *Interface and Interface Conditions. Language, Context and Cognition*, ed A. Späth (Berlin; New York: de Gruyter), 277–297.
- Hoehle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., and Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: evidence from German and French infants. *Infant Behav. Dev.* 32, 262–274. doi: 10.1016/j.infbeh.2009.03.004
- Huijbers, W., Pennartz, C. M. A., Rubin, D. C., and Daselaar, S. M. (2011). Imagery and retrieval of auditory and visual information: neural correlates of successful and unsuccessful performance. *Neuropsychologia* 49, 1730–1740. doi: 10.1016/j.neuropsychologia.2011.02.051
- Ischebeck, A., Friederici, A. D., and Alter, K. (2008). Processing prosodic boundaries in natural and hummed speech. An fMRI study. *Cereb. Cortex* 18, 541–552. doi: 10.1093/cercor/bhm083
- Janssen, U., and Domahs, F. (2008). Going on with optimized feet: evidence for the interaction between segmental and metrical structure from a case of Primary Progressive Aphasia. *Aphasiology* 22, 1157–1175. doi: 10.1080/02687030701820436
- Jessen, M., and Marasek, K. (1997). “Voice quality correlates of word stress in tense versus lax vowels in German,” in *Larynx* (Marseille), 127–130.
- Just, M. A., Newman, S. D., Keller, T. A., McEleny, A., and Carpenter, P. (2004). Imagery insentence comprehension: an fMRI study. *Neuroimage* 21, 112–124. doi: 10.1016/j.neuroimage.2003.08.042
- Kaltenbacher, E. (1994). Typologische Aspekte des Wortakzents: Zum Zusammenhang von Akzentposition und Silbengewicht im Arabischen und im Deutschen. *Zeitschrift für Sprachwissenschaft* 13, 20–55. doi: 10.1515/zfsw.1994.13.1.20
- Klein, E., Domahs, U., Grande, M., and Domahs, F. (2011). Neuro-cognitive foundations of word stress processing—evidence from fMRI. *Behav. Brain Funct.* 7:15. doi: 10.1186/1744-9081-7-15
- Klein, E., Willmes, K., Dressel, K., Domahs, F., Wood, G., Nuerk, H.-C., et al. (2010). Categorical and continuous—disentangling the neural correlates of the carry effect in multi-digit addition. *Behav. Brain Funct.* 6:70. doi: 10.1186/1744-9081-6-70

- Knaus, J., and Domahs, U. (2009). Experimental evidence for optimal and minimal metrical structure of German word prosody. *Lingua* 119, 1396–1413. doi: 10.1016/j.lingua.2008.04.002
- Kong, J., Wang, C., Kwong, K., Vangel, M., Chuac, E., and Gollub, R. (2005). The neural substrate of arithmetic operations and procedure complexity. *Cogn. Brain Res.* 22, 397–405. doi: 10.1016/j.cogbrainres.2004.09.011
- Kotz, S. A., and Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn. Sci.* 14, 392–399. doi: 10.1016/j.tics.2010.06.005
- Kotz, S. A., and Schwartze, M. (2011). Differential input of the supplementary motor area to a dedicated temporal processing network: functional and clinical implications. *Front. Integr. Neurosci.* 5, 1–4. doi: 10.3389/fnint.2011.00086
- Kotz, S. A., Schwartze, M., and Schmidt-Kassow, M. (2009). Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. *Cortex* 45, 982–990. doi: 10.1016/j.cortex.2009.02.010
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–75. doi: 10.1017/S0140525X99001776
- Lewis, J. W., Beauchamp, M. S., and DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb. Cortex* 10, 873–888. doi: 10.1093/cercor/10.9.873
- Lewis, P. A., and Miall, R. C. (2003). Brain activation patterns during measurement of sub- and supra-second intervals. *Neuropsychologia* 41, 1583–1592. doi: 10.1016/S0028-3932(03)00118-0
- Lewis, P. A., Wing, A. M., Pope, P. A., Praamstra, P., and Miall, R. C. (2004). Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. *Neuropsychologia* 42, 1301–1312. doi: 10.1016/j.neuropsychologia.2004.03.001
- Li, X., Gandour, J. T., Talavage, T., Wong, D., Hoffa, A., Lowe, M., et al. (2010). Hemispheric asymmetries in phonological processing of tones versus segmental units. *Neuroreport* 21, 690–694. doi: 10.1097/WNR.0b013e32833b0a10
- Lintfert, B. (2010). *Phonetic and Phonological Development of Stress in German*. Doctoral thesis, Universität Stuttgart, Stuttgart.
- Lutz, K., Specht, K., Shah, N. J., and Jancke, L. (2000). Tapping movements according to regular and irregular visual timing signals investigated with fMRI. *Neuroreport* 11, 1301–1306. doi: 10.1097/00001756-200004270-00031
- Mehler, J., Dupoux, E., Pallier, C., and Dehaene-Lambertz, G. (2004). Cross-linguistic approaches to speech processing. *Curr. Opin. Neurobiol.* 4, 171–176. doi: 10.1016/0959-4388(94)90068-X
- Molczanow, J., Domahs, U., Knaus, J., and Wiese, R. (2013). The lexical representation of word stress in Russian: evidence from event-related potentials. *Ment. Lex.* 8, 164–194. doi: 10.1075/ml.8.2.03mol
- Peperkamp, S., and Dupoux, E. (2002). “A typological study of stress “deafness,” in *Laboratory Phonology 7*, eds C. Gussenhoven, and N. Warner (Berlin: Walter de Gruyter GmbH & Co. KG), 203–240.
- Peperkamp, S., Dupoux, E., and Sebastián-Gallés, N. (1999). “Perception of stress by French, Spanish, and bilingual subjects,” in *ESCA 7th European Conference on Speech Communication and Technology* (Budapest).
- Peperkamp, S., Vendelin, I., and Dupoux, E. (2010). Perception of predictable stress: a cross-linguistic investigation. *J. Phon.* 38, 422–430. doi: 10.1016/j.wocn.2010.04.001
- Riecker, A., Kassubek, J., Gröschel, K., Grodd, W., and Ackermann, H. (2006). The cerebral control of speech tempo: opposite relationship between speaking rate and BOLD signal changes at striatal and cerebellar structures. *Neuroimage* 29, 46–53. doi: 10.1016/j.neuroimage.2005.03.046
- Roettger, T., Domahs, U., Grande, M., and Domahs, F. (2012). Structural factors affecting the assignment of word stress in German. *J. Ger. Linguist.* 24, 53–94. doi: 10.1017/S1470542711000262
- Sadato, N., Campbell, G., Ibáñez, V., Deiber, M., and Hallett, M. (1996). Complexity affects regional cerebral blood flow change during sequential finger movements. *J. Neurosci.* 16, 2691–2700.
- Sakai, K., Hikosaka, O., Miyachi, S., Takino, R., Sasaki, Y., and Pütz, B. (1998). Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. *J. Neurosci.* 18, 1827–1840.
- Sakurai, Y. (1990). Hippocampal cells have behavioral correlates during the performance of an auditory working memory task in the rat. *Behav. Neurosci.* 104, 253–263. doi: 10.1037/0735-7044.104.2.253
- Sakurai, Y. (1994). Involvement of auditory cortical and hippocampal neurons in auditory working memory and reference memory in the rat. *J. Neurosci.* 14, 2606–2623.
- Scheperjans, F., Eickhoff, S. B., Hömke, L., Mohlberg, L., Hermann, K., Amunts, K., et al. (2008a). Probabilistic maps, morphometry, and variability of cytoarchitectonic areas in the human superior parietal cortex. *Cereb. Cortex* 18, 2141–2157. doi: 10.1093/cercor/bhm241
- Scheperjans, F., Hermann, K., Eickhoff, S. B., Amunts, K., Schleicher, A., and Zilles, K. (2008b). Observer-Independent Cytoarchitectonic Mapping of the Human Superior Parietal Cortex. *Cereb. Cortex* 18, 846–867. doi: 10.1093/cercor/bhm116
- Schmidt-Kassow, M., Roncaglia-Denissen, M., and Kotz, S. A. (2011b). Why pitch sensitivity matters: event-related potential evidence of metric and syntactic violation detection among Spanish late learners of German. *Front. Psychol.* 2, 1–11. doi: 10.3389/fpsyg.2011.00131
- Schmidt-Kassow, M., Rothermich, K., Schwartze, M., and Kotz, S. A. (2011a). Did you get the beat? Late proficient French-German learners extract strong-weak patterns in tonal but not in linguistic sequences. *Neuroimage* 54, 568–576. doi: 10.1016/j.neuroimage.2010.07.062
- Schneider, K. (2007). “Acquisition of word stress in German: vowel duration and incompleteness of closure,” in *Proceedings of the ICPHS XVI* (Saarbrücken), 1565–1568.
- Schneider, K., and Möbius, B. (2007). “Word stress correlates in spontaneous child-directed speech in German,” in *Proceedings of the Interspeech* (Antwerp), 1394–1397.
- Schwartz, M., Rothermilch, K., and Kotz, S. A. (2012a). Functional dissociation of pre-SMA and SMA-proper in temporal processing. *Neuroimage* 60, 290–298. doi: 10.1016/j.neuroimage.2011.11.089
- Schwartz, M., Tavano, A., Schröger, E., and Kotz, S. A. (2012b). Temporal aspects of prediction in audition: cortical and subcortical neural mechanisms. *Int. J. Psychophysiol.* 83, 200–207. doi: 10.1016/j.ijpsycho.2011.11.003
- Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406. doi: 10.1093/brain/123.12.2400
- Shibasaki, H., Sadato, N., Lyshkow, H., Yonekura, Y., Honda, M., Nagamine, T., et al. (1993). Both primary motor cortex and supplementary motor area play an important role in complex finger movement. *Brain* 116, 1387–1398. doi: 10.1093/brain/116.6.1387
- Stanton, H. M. (1928). Seashore measures of musical talent. *Psychol. Monogr. Gen. Appl.* 39, 135–144. doi: 10.1037/h0093342
- Tappeiner, E., Domahs, U., and Domahs, F. (2007). Wortakzent im Sprachkontext Deutsch-Italienisch. *Zeitschrift für Dialektologie und Linguistik* 74, 266–291.
- Tewes, U. (1991). *HAWIE-R. Hamburg-Wechsler Intelligenstest für Erwachsene. Revision*. Bern: Huber.
- Van der Hulst, H. (ed.). (1999). *Word Prosodic Systems in the Languages of Europe. Empirical Approaches to Language Typology*, Vol. 4. Berlin; New York: Walter de Gruyter.
- Vennemann, T. (1990). Syllable structure and simplex accent in modern standard German. *CLS* 26, 399–412.
- Vennemann, T. (1991). “Syllable structure and simplex accent in Modern Standard German,” in *Papers from the 26th Regional Meeting of the Chicago Linguistic Society. II. The Parasession on the Syllable in Phonetics and Phonology* (CLS 26), eds M. Ziolkowski, M. Noske, and K. Deaton (Chicago, IL: Chicago Linguistic Society), 399–412.
- Wiese, R. (1996). *The Phonology of German*. Oxford: Clarendon Press.
- Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., et al. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cereb. Cortex* 14, 1384–1389. doi: 10.1093/cercor/bhh099
- Wood, G., Ischebeck, A., Koppelstaetter, F., Gotwald, T., and Kaufmann, L. (2009). Developmental trajectories of magnitude processing and interference control: an fMRI study. *Cereb. Cortex* 19, 2755–2765. doi: 10.1093/cercor/bhp056
- Yonelinas, A. P. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. *Behav. Brain Res.* 254, 34–44. doi: 10.1016/j.bbr.2013.05.030

- Zarate, J. M., and Zatorre, R. J. (2005). Neural substrates governing audiovocal integration for vocal pitch regulation in singing. *Ann. N.Y. Acad. Sci.* 1060, 404–408. doi: 10.1196/annals.1360.058
- Zilles, K., and Rehämper, G. (1998). *Funktionelle Neuroanatomie* (3. komplett überarbeitete und aktualisierte Auflage). Berlin: Springer-Verlag.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 January 2014; accepted: 07 April 2014; published online: 29 April 2014.

*Citation:* Heisterueber M, Klein E, Willmes K, Heim S and Domahs F (2014) Processing word prosody—behavioral and neuroimaging evidence for heterogeneous performance in a language with variable stress. *Front. Psychol.* 5:365. doi: 10.3389/fpsyg.2014.00365

This article was submitted to Language Sciences, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Heisterueber, Klein, Willmes, Heim and Domahs. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Phonetic detail in German syllable pronunciation: influences of prosody and grammar

Barbara Samlowski<sup>1\*</sup>, Bernd Möbius<sup>2</sup> and Petra Wagner<sup>1</sup>

<sup>1</sup> Work Group Phonetics/Phonology, Faculty of Linguistics and Literary Studies, Bielefeld University, Bielefeld, Germany

<sup>2</sup> Department of Computational Linguistics and Phonetics, Saarland University, Saarbrücken, Germany

## Edited by:

Richard Wiese, Philipps-Universität Marburg, Germany

## Reviewed by:

Christiane Ulbrich, University of Marburg, Germany  
Marzena Zygis, Centre for General Research & Humboldt University, Germany

## \*Correspondence:

Barbara Samlowski, Work Group Phonetics/Phonology, Faculty of Linguistics and Literary Studies, Bielefeld University, Universitätsstraße 25, Bielefeld 33615, Germany  
e-mail: barbara.samlowski@uni-bielefeld.de

This study presents two experiments designed to disentangle various influences on syllable pronunciation. Target syllables were embedded in carrier sentences, read aloud by native German participants, and analyzed in terms of syllable and vowel duration, acoustic prominence, and spectral similarity. Both experiments revealed a complex interaction of different factors, as participants attempted to disambiguate semantically and syntactically ambiguous structures while at the same time distinguishing between important and unimportant information. The first experiment examined German verb prefixes that formed prosodic minimal pairs. Carrier sentences were formulated so as to systematically vary word stress, sentence focus, and the type of syntactic boundary following the prefix. We found clear effects of word stress on duration, prominence, and spectral similarity as well as a small influence of sentence focus on prominence levels of lexically stressed prefixes. While sentence boundaries were marked by particularly high prominence and duration values, hardly any effect was shown for word boundaries. The second experiment compared German function words which were segmentally identical but appeared in different grammatical roles. Here, definite articles were found to be shorter than relative pronouns and still shorter than demonstrative pronouns. As definite articles are also much more common than the other two lexical classes, effects of lemma frequency might also have played a role.

**Keywords:** prominence, duration, stress, syntactic boundaries, lexical class, lemma frequency

## INTRODUCTION

Syllables can vary strongly in the way they are pronounced, even when in canonical pronunciation they are segmentally identical. One important source of variation is prominence, i.e., the degree of emphasis which is placed on syllables and with which they are perceived. Such emphasis may be realized by means of higher duration and intensity values, overall larger articulatory effort, as well as the presence and shape of pitch accents (Wagner, 2002). Among other things, prominence differences are used to distinguish between lexically stressed and unstressed syllables. Duration seems to be a main correlate of word stress in German, but differences were also found for formant values, fundamental frequency, and various voice quality parameters (e.g., Kohler, 1987; Claßen et al., 1998; Kleber and Klippahhn, 2006; Schneider and Möbius, 2007; Lintfert, 2010). Studies specifically investigating word stress in focused and unfocused sentence positions have confirmed duration as a strong signal of word stress which operates independently of sentence accent (Dogil and Williams, 1999 for German; Okobi, 2006 and Cho and Keating, 2009 for English; Sluijter and van Heuven, 1996 for Dutch). However, for English, Plag et al. (2011) found no effect at all of word stress on duration, while Campbell and Beckman (1997) discovered stress-related duration differences only in one of the two unaccented contexts examined. For English and Dutch, spectral tilt, i.e., the intensity in higher compared to lower frequency bands, appeared to be another robust

correlate of word stress in accented as well as unaccented contexts (Sluijter and van Heuven, 1996; Okobi, 2006; Plag et al., 2011). Although Dogil and Williams (1999) found no significant differences between accented and unaccented words in German in terms of fundamental frequency, intensity, or duration, studies for other languages showed stress-related differences in fundamental frequency and intensity to be strongly reduced when target words were not accented (Sluijter and van Heuven, 1996; Plag et al., 2011). Apart from signaling sentence focus, prominence differences are also used to distinguish important from unimportant information on the level of lexical class. In German speech synthesis, lexical class has been used as an important indicator for predicting prominence levels (Widera et al., 1997; Windmann et al., 2011). Frequency and predictability effects have an influence on word pronunciation as well. There is evidence for English that words tend to be spoken at a faster rate if they are frequent or easily predictable from their context (Bell et al., 2003; Aylett and Turk, 2004; Baker and Bradlow, 2009). Although effects of word frequency and lexical class are often confounded, both factors were found to play an important role (Jurafsky et al., 2000; Pluymaekers et al., 2005; Bell et al., 2009). The present study consists of two controlled production experiments. The first experiment aims to disentangle influences of lexical stress, sentence accent, and syntactic boundaries, while the second experiment analyzes effects of lexical class and word frequency.



## METHODS

### PARTICIPANTS

Thirty participants took part in the two experiments (15 men, 15 women, ages ranging between 19 and 47). All were native speakers of German. They were paid for their participation in the study.

### MATERIAL

#### *Experiment 1: Stress, accent, and syntactic boundaries*

Certain German verbs can differ in meaning depending on whether their lexical stress falls on the prefix or the verb stem. For example, the word [ʔʊn.tə.ʃtɛ.lən] ([unter]<sub>prefix</sub>–[stɛll]<sub>stem</sub>–[ən]<sub>ending</sub>verb), literally “to under-put” means “to store / take shelter” when stressed on the prefix, but “to insinuate” when stressed on the stem. This ambiguity is not visible in all inflections, however. In most finite forms, lexically stressed prefixes are separated from the verb and placed at the end of the clause. As the verb prefixes used for this experiment are segmentally identical to prepositions or conjunctions, we were able to use them to analyze effects of syntactic boundaries as well. We examined the effects of word and sentence stress as well as word and sentence boundaries on the production of the four German verb prefixes “um” ([ʔʊm] – “around”), “unter” ([ʔʊn.tə] – “under”), “über” ([ʔy:.bɐ] – “over”), and “durch” ([dʊʁç] – “through”) in a reading task (see also Samlowski et al., 2012). The phonetic transcriptions given here are canonical. The glottal stop preceding onset vowels may be omitted or realized through vowel glottalization, and the [ʁ] in “durch” is commonly rendered as [ɐ].

Target items consisted of the prefixes combined with two different verb stems each. Each of the eight resulting verbs was placed in seven different carrier sentences. In sentences 1–4, word and sentence stress were varied, while sentences 5–7 compared different types of syntactic boundaries (see **Table 1**). As participants needed to be able to infer the correct stress pattern from the sentence context, a different set of carrier sentences was created for each verb. Sentence stress differences were not elicited in a uniform manner, either. While sentences belonging to the categories “w+s+” and “w-s+” were formulated so as to imply a broad focus, sentences in categories “w+s-” and “w-s-” contained elements designed to attract a contrasting focus and thereby move the sentence stress away from the main verb. Among the strategies used for this were the inclusion of two contrasting objects, topic fronting, and the addition of an emphasized modifier. For the sake of brevity in this paper we refer to the first four sentence categories in terms of stressed and unstressed prefixes (“w+” vs. “w-”) in accented and unaccented conditions (“s+” vs. “s-”). Nonetheless, it is important to note that the categories do not reflect the actual stress patterns used by the participants. Instead they describe potential differences in word and sentence stress due to different word meanings and the presence or absence of an additional motivation for deaccentuating the verb. Our aim is to discover the extent to which these conceptual differences are realized in the acoustic production of the target syllables.

We deliberately decided against using underline, font style, or a question-answer structure to indicate lexical stress and sentence focus, since we wanted to avoid potentially evoking exaggerated responses by attracting the participants’ attention to the intended

**Table 1 | Sentence categories.**

Category	Canonical word stress	Additional semantic contrast	Right-hand boundary
w+s+	<b>Yes</b>	<b>No</b>	(morpheme)
w+s-	<b>Yes</b>	<b>Yes</b>	(morpheme)
w-s+	<b>No</b>	<b>No</b>	(morpheme)
w-s-	<b>No</b>	<b>Yes</b>	(morpheme)
sb	(undefined)	(no)	<b>Sentence</b>
mb	(no)	(no)	<b>Morpheme</b>
wb	(undefined)	(no)	<b>Word</b>

*Word and sentence stress status and succeeding syntactic boundary for the target items in each of the 7 sentences (manipulated parameters in bold).*

reading. This meant that the context was not controlled across verbs and only up to a limited degree within each set of sentences. For the first four sentences within one set, the half syllable preceding and following the prefix were kept constant. Sentence 5 (“sb”) used the same preceding half-syllable as the first four. While the prefix in sentence 6 (“mb”) fulfilled the same conditions as in sentence 3 (“w-s+”), its target sentence was formulated so that the preceding and following half-syllables matched those of the identical prepositions or conjunctions in sentence 7 (“wb”).

#### *Experiment 2: Lexical class and word frequency*

While different words are used for German demonstrative pronouns, relative pronouns, and definite articles, depending on gender, number, and case, these words are often segmentally identical across the three lexical classes. Definite articles are much more common than the segmentally identical demonstrative or relative pronouns. According to the DeWaC corpus (Baroni and Kilgarriff, 2006), a 1.5 billion word database of German internet articles which was automatically tagged for lexical classes, the words “der” ([de:r]), “die” ([di:]), “das” ([das]), “dem” ([de:m]), and “den” ([de:n]) were used as definite articles 89.8% of the time, while 7.2% of their appearances were classified as relative pronouns, and only 3% were demonstrative pronouns.

To examine whether these differences in frequency of occurrence have an influence on pronunciation, we compared their realizations in different grammatical roles (see also Samlowski et al., 2013). Sentences containing relative and demonstrative pronouns were formulated so as to match definite articles already appearing in one of the other carrier sentences from the two experiments. As each of the lexical classes required different types of surrounding grammatical structure, only the half-syllables preceding and following the target word were held constant across each group of 3 sentences. For each of the investigated words, 3 sentence groups were assembled (see **Table 2**), resulting in a total of 48 new sentences containing relative and demonstrative pronouns.

### PROCEDURE

Sentences from both experiments were placed in a quasi-random order, which was not varied across participants. Care was taken to avoid repetitions of the same verb and provide a good mixture of sentences from both experiments, allowing them to function as mutual distractors. Acoustic recordings took place in a



**Table 2 | Target items.**

Orthographic form	Phonetic transcription	Gender / number / case	No. of stimuli
der	[de:rə]	<b>Masculine</b> singular nominative <b>or</b> <b>Feminine</b> singular dative	3 × 3 3 × 3
die	[di:]	Feminine <b>singular</b> nominative/accusative <b>or</b> Masculine/feminine/neuter <b>plural</b> nominative/accusative	3 × 3 3 × 3
das	[das]	Neuter singular nominative/accusative	3 × 3
dem	[de:m]	<b>Masculine</b> singular dative <b>or</b> <b>Neuter</b> singular dative	3 × 3 3 × 3
den	[de:n]	Masculine singular accusative	3 × 3

Orthographic form, phonetic transcription, grammatical description and number of stimuli used for each word.

sound-treated chamber at Bielefeld University. One sentence at a time was presented on a computer screen to the participants, who proceeded through the experiment in a self-paced manner. To further clarify the intended word meaning and improve understanding of the reading content, sentences were illustrated using the text-to-scene conversion program WordsEye (Coyne and Sproat, 2001, see **Figures 1** and **2**). Participants looked at each sentence and the accompanying picture and then read the sentence out loud. Beforehand, they were instructed to repeat any sentences in which they made a mistake or slip of the tongue. These sentences as well as sentences where participants hesitated noticeably while reading were omitted from analysis. Target items were also discarded if they or their immediate context was impaired through speech errors, noise, or unexpected vowel elision.

The remaining recordings were analyzed in terms of syllable and vowel duration, acoustic prominence, and spectral similarity. For the duration and prominence analysis, syllable and vowel boundaries of the target items as well as the preceding and following syllables were manually annotated with Praat (Boersma, 2001). Acoustic prominence was investigated by means of an automatic prominence tagger which analyzed annotated syllable nuclei in terms of pitch movement, duration, intensity, and spectral emphasis. Values for the last three parameters were normalized across all investigated syllables in the utterance using z-scores and the individual factors were weighted so as to model perceptual ratings of German prominence (Tamburini and Wagner, 2007). In the present study, only the syllables immediately preceding or following the target items were used as context for the tagger. If the vowel of a context syllable tended to be elided, the preceding/following syllable nucleus was taken as context syllable instead. We also compared pairs of segmentally identical syllables produced by the same speaker in terms of spectral similarity, using a method developed by Wade and Möbius (2007) and Lewandowski (2011). Amplitude envelopes were computed for 4 frequency bands (equally spaced on a logarithmic scale ranging



**FIGURE 1 | Example illustration for Experiment 1—“unterstellen” (category “w+s+”).** Corresponding sentence: “Wir wollten uns **unterstellen**, weil es so stark regnet.” (English: “We wanted to **take shelter** because it is raining so heavily.”)



**FIGURE 2 | Example illustration for Experiment 2—“den” (rp).** Corresponding sentence: “Es war deutlich, dass der Fuchs den See beobachtete, **den** Enten als ihre Heimat gewählt hatten.” (English: “It was clear that the fox was watching the lake **which** ducks had chosen as their home.”)

from 80 to 7800 Hz), using a sampling rate of 500 Hz. The spectral similarity of two syllables was calculated by cross-correlating pairs of envelopes for each frequency band, taking the maximum of the cross-correlation as an indicator for the degree of similarity. Although spectral similarity is not a direct measure of vowel quality and degree of coarticulation, it can serve as an indication of how strongly the target items varied in their pronunciation across contexts and categories. Statistical analysis and visualization was performed with R (R Development Core Team, 2010). As residuals from analyses of variances only followed a normal distribution in the case of the duration results of the second experiment, the other investigations were analyzed with Wilcoxon rank sum tests. Significance values were Bonferroni-corrected for multiple comparisons.

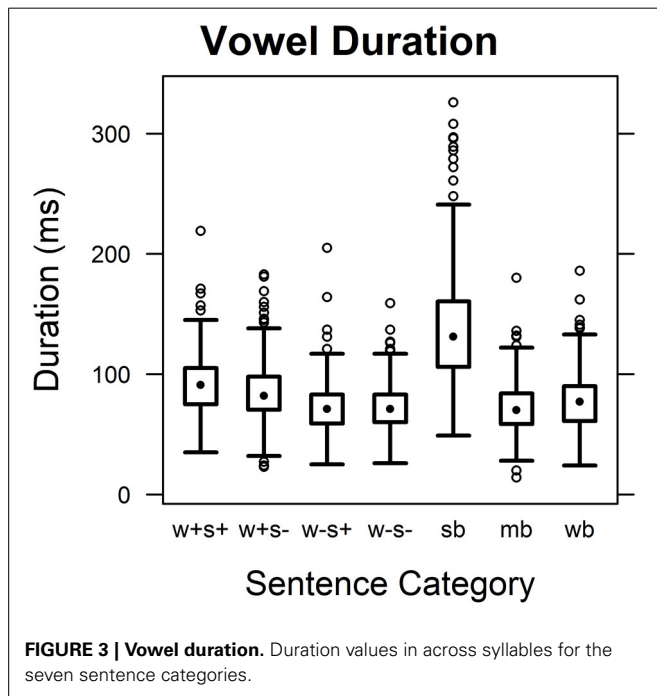
## RESULTS

### EXPERIMENT 1: STRESS, ACCENT, AND SYNTACTIC BOUNDARIES

Of the 1680 sentences collected (8 verbs  $\times$  7 sentences  $\times$  30 participants), 113 were discarded. As two of the prefixes used are bisyllabic, the following analyses are based on a total of 2278 syllables. The results were analyzed in terms of sentence category (“w+s+,” “w+s-,” “w-s+,” “w-s-,” “sb,” “mb,” “wb”) and syllable identity ([ʔum], [ʔun], [tə], [ʔy:], [bɐ], [dʊɐç]).

#### Duration

**Figure 3** gives an overview of vowel duration results for the seven sentence categories examined. Wilcoxon rank sum tests



comparing sentence categories across syllables (corrected for 21 comparisons) showed phrase-final prefixes (“sb”) to be significantly longer than those in the other categories ( $W > 87,000$ ,  $p < 0.0001$ ). A small influence of word stress was also observed, with syllables and vowels being longer when appearing in lexically stressed compared to unstressed prefixes (“w+s+” vs. “w-s+,” “w+s-” vs. “w-s-,”  $W > 62,000$ ,  $p < 0.0001$ ). Vowel duration of lexically stressed prefixes was slightly reduced if the verb was not in the focus of the sentence (“w+s+” vs. “w+s-,”  $W = 59,074$ ,  $p < 0.05$ ). Finally, there was a small tendency for prepositions or conjunctions to have slightly longer syllables and vowels than segmentally identical bound prefixes (“wb” vs. “mb,”  $W > 59,000$ ,  $p < 0.05$ ).

Syllable and vowel durations were also analyzed for combinations of syllable identity and sentence category (corrected for 861 comparisons, see **Table 3** for mean values). All investigated syllables were significantly longer when they occurred in separated sentence-final prefixes than in other contexts (“sb” vs. others,  $W > 2400$ ,  $p < 0.0001$ ). Differences in vowel duration were significant for all syllables except [ʔun]. Here, differences between separated prefixes and bound prefixes in lexically stressed and potentially accented positions (“sb” vs. “w+s+”) failed to reach significance, and comparisons between separated prefixes and segmentally identical function words (“sb” vs. “wb”) were significant on a lower level ( $W = 2386$ ,  $p < 0.01$ ) than the other comparisons ( $W > 2200$ ,  $p < 0.0001$ ). No significant influences were shown for word boundary (“mb” vs. “wb”) or sentence stress (“w+s+” vs. “w+s-,” “w-s+” vs. “w-s-”). Effects of word stress on syllable and vowel duration are summarized in **Table 4**.

#### Prominence

Prominence estimates for the individual syllables in the seven sentence categories are shown in **Figure 4**. Wilcoxon rank sum tests for sentence categories across syllables (corrected for 21 comparisons) showed that lexically stressed prefixes tended to receive significantly higher prominence values than unstressed ones in accented as well as unaccented conditions (“w+s+” vs.

**Table 3 | Mean duration.**

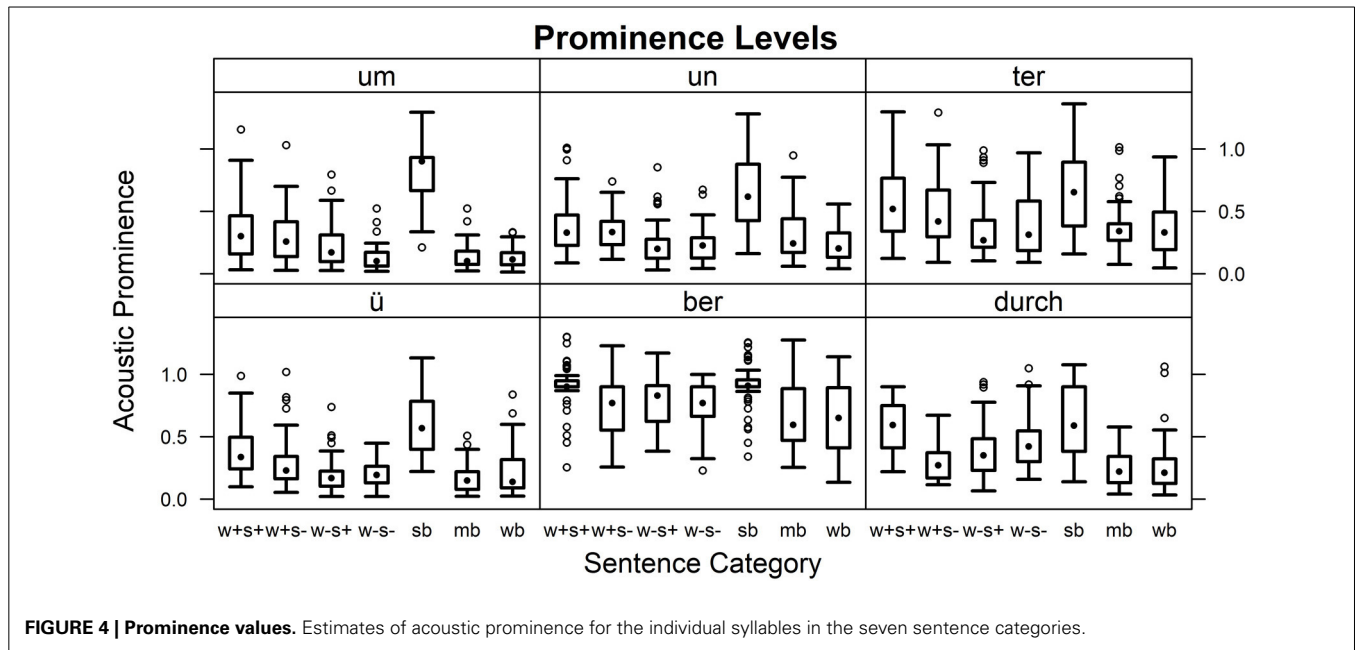
Category	[ʔum]	[ʔun]	[tə]	[ʔy:]	[bɐ]	[dʊɐç]	All
w+s+	147.6	142.4	132.1	90.0	138.0	199.1	140.0
	72.2	88.6	88.0	90.0	87.5	114.4	91.7
w+s-	151.4	130.7	125.6	84.8	136.6	194.7	137.0
	77.1	79.4	79.2	84.8	85.5	110.2	85.6
w-s+	143.4	112.6	115.5	59.9	126.3	149.0	117.0
	71.6	66.1	74.7	59.9	75.7	81.2	71.3
w-s-	140.5	112.5	114.5	66.9	127.0	161.9	120.7
	66.5	65.5	73.0	66.9	77.2	88.7	72.9
sb	261.0	185.2	205.5	132.3	207.1	414.2	228.7
	128.6	106.3	137.8	132.3	147.1	188.0	140.1
mb	127.0	116.3	119.5	59.6	121.9	146.9	114.5
	61.7	71.1	76.4	59.6	74.1	83.5	71.2
wb	140.2	127.8	124.5	71.4	127.8	152.8	124.4
	71.9	82.5	76.5	71.4	76.9	84.3	77.3

Mean duration values in milliseconds for syllables (above) and vowels (below) in the seven sentence categories.

**Table 4 | Duration statistics for lexical stress.**

Category	[ʔum]	[ʔun]	[tə]	[ʔy:]	[bə]	[dʊɐ̯ç]
w+s+ vs. w-s+	1868.0 (n.s.) 1756.5 (n.s.)	2485.0**** 2423.5****	2214.5* 2177.5 (n.s.)	2687**** 2687****	2264.5 (n.s.) 2365.5**	2063.5**** 2199****
w+s- vs. w-s-	1743.0 (n.s.) 1870.5 (n.s.)	2353.5* 2309.5*	2150.0 (n.s.) 2116.5 (n.s.)	1921* 1921*	1670.5 (n.s.) 1674 (n.s.)	1925.5 (n.s.) 1945.4*

W values with significance levels for syllables (above) and vowels (below), n.s.:  $p \geq 0.05$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\*\* $p < 0.0001$ .

**FIGURE 4 | Prominence values.** Estimates of acoustic prominence for the individual syllables in the seven sentence categories.

“w-s+,”  $W = 72,651$ ,  $p < 0.0001$ ; “w+s-” vs. “w-s-,”  $W = 59,677$ ,  $p < 0.01$ ). Sentence stress differences were significant for lexically stressed prefixes (“w+s+” vs. “w+s-,”  $W = 63,503$ ,  $p < 0.0001$ ). Separated, phrase-final prefixes were particularly high in prominence (“sb” vs. others,  $W > 76,000$ ,  $p < 0.0001$ ), while no effect of word boundary was observed (“mb” vs. “wb”).

In tests for combinations of sentence categories and syllables (corrected for 821 comparisons) differences related to word and sentence stress mostly failed to reach significance. Word stress effects were found for [ʔy:] and [dʊɐ̯ç] in accented conditions as well as for [ʔum] and [ʔun] in unaccented conditions (see **Table 5**). Effects of sentence stress were only shown in the case of lexically stressed [dʊɐ̯ç] (“w+s+” vs. “w+s-,”  $W = 2189$ ,  $p < 0.0001$ ). In separated, phrase-final prefixes, syllables often received significantly higher prominence values than in the other categories (“sb” vs. others,  $W > 1900$ ,  $p < 0.05$ ). Exceptions for this last tendency were found for [tə] (“sb” vs. “w+s+,” “sb” vs. “w+s-”), [bə] (“sb” vs. “w+s+,” “sb” vs. “w-s+”), and [dʊɐ̯ç] (“sb” vs. “w+s+,” “sb” vs. “w-s-”). No significant differences appeared between bound prefixes and corresponding prepositions or conjunctions (“mb” vs. “wb”) or between unstressed prefixes in accented and unaccented conditions (“w-s+” vs. “w-s-”).

### Spectral similarity

For each target syllable in each sentence category, we calculated the level of similarity between prefixes produced by the same speaker in the two verb contexts. Wilcoxon rank sum tests comparing sentence categories across syllables (corrected for 21 comparisons) showed significant differences in syllable similarity for stressed versus unstressed prefixes in accented conditions (“w+s+” vs. “w-s+,”  $W = 16,183.5$ ,  $p < 0.01$ , mean values: 0.889 vs. 0.848). Sentence stress differences in stressed prefixes had only a marginally significant effect (“w+s+” vs. “w+s-,”  $W = 14,189.5$ ,  $p = 0.052$ , mean values: 0.889 vs. 0.864). Separated, phrase-final prefixes (“sb,” mean: 0.897) received significantly higher similarity values ( $W > 16,000$ ,  $p < 0.001$ ) compared to all examined categories except for stressed and potentially accented prefixes (“w+s+”). Effects were most pronounced for the syllables [ʔy:], [dʊɐ̯ç], and, to a lesser extent, [ʔun], although results failed to reach significance when combinations of syllables and sentence categories were investigated (corrected for 821 comparisons).

In an analysis of spectral similarity between sentence categories for syllables produced by the same speaker in the same verb context, comparisons with separated, phrase-final prefixes tended to result in lower values than comparisons between the other

**Table 5 | Prominence statistics for lexical stress.**

Categories	[ʔum]	[ʔun]	[tə]	[ʔy:]	[bə]	[dʊʁç]
w+s+ vs. w-s+	1922.5 (n.s.)	2171.5 (n.s.)	2174 (n.s.)	2669.5****	2085 (n.s.)	1977**
w+s- vs. w-s-	2357.5****	2307.5*	1985 (n.s.)	1664.5 (n.s.)	1240.5 (n.s.)	741 (n.s.)

*W* values with significant levels for comparisons between lexically stressed and unstressed prefixes, n.s.:  $p \geq 0.05$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\*\* $p < 0.0001$ .

sentence categories (“sb” vs. others). This effect was shown to be significant ( $W > 57,000$ ,  $p < 0.0001$ ) in tests for combinations of sentence categories (corrected for 210 comparisons). Lexically stressed prefixes were significantly closer to those in sentence-final prefixes than syllables in unstressed prefixes (“w+s+” and “sb” vs. “w-s+” and “sb,” “w+s-” and “sb” vs. “w-s-” and “sb,”  $W > 64,000$ ,  $p < 0.001$ , mean values: 0.794 vs. 0.748 and 0.799 vs. 0.765). Here as well as for the comparisons within sentence categories, effects were most clearly visible for [ʔy:] and [dʊʁç]. An analysis of similarity between sentences in the “sb” category and those in the other categories combined with syllable identity (corrected for 630 comparisons) showed significant differences between stressed and unstressed [ʔy:] in accented conditions (“w+s+” and “sb” vs. “w-s+” and “sb,” mean values: 0.816 vs. 0.709,  $W = 2552$ ,  $p < 0.0001$ ).

### Discussion

Apart from Dogil and Williams (1999), there have been hardly any studies examining the interaction of word and sentence stress in German. In our paper, we examine the extent to which canonical word stress differences and additional semantic contrasts triggered differences in the word and sentence stress patterns which in turn were visible in the acoustic realization of the target syllables. Based on German language corpus studies as well as evidence from other Germanic languages, we expected lexically stressed syllables to be longer than unstressed syllables in accented as well as unaccented conditions. We also predicted an effect of word and sentence stress on acoustic prominence levels compared to the immediate surroundings. Although spectral parameters have been shown to be affected by stress, we had no clear hypotheses as to how word and sentence stress might influence similarity across and within sentence categories. Our study indeed showed a significant influence of lexical stress on duration values for all investigated prefixes apart from [ʔum]. When sentences were given a broad focus, even the lexically unstressed second syllables of the prefixes [ʔun.tə] and [ʔy:.bə] were affected. This result may be explained by accentual lengthening of the word carrying sentence stress, as there is evidence that in English and Dutch this effect is stronger to the right of the lexically stressed syllable than to the left (Cambier-Langeveld and Turk, 1999). There was also a tendency for stressed syllables to be higher in prominence and more similar to syllables in sentence-final prefixes than unstressed ones. When no deaccentuation cues were given, lexically stressed syllables were more similar across verb contexts than unstressed syllables. Results for prominence and spectral similarity mostly failed to reach significance in a syllable-by-syllable analysis. One reason for the small size of the word stress effects might be that all investigated syllables except [ʔy:]

had lax vowels, since these have been found to have a considerably reduced effect of lexical stress on duration (Mooshammer et al., 1999; Kleber and Klippahhn, 2006). Although there was a slight effect of sentence stress on duration and prominence values of lexically stressed syllables, it almost never reached significance in a syllable-by-syllable analysis. Although the data was not analyzed perceptually, auditory impressions suggest that participants often placed a secondary accent on the target verb in unaccented conditions—perhaps because they wanted to better clarify the intended word meaning or because the given cues were not strong enough. Particularly in the case of the verbs [ʔdʊʁç.fəʊ.ən] (“to look through”) and [ʔʊm.fa:.ʁən] (“to run over”), effects of final lengthening might also have played a role, as these were sentence-final in the unaccented, but not in the accented conditions. The unusually strong effect of sentence stress on prominence levels for [dʊʁç] may have been due to the fact that [ʔdʊʁç.fəʊ.ən] was one of the few verbs where the potentially contrasting sentence stress in the unaccented condition would actually fall on the syllable used as preceding context by the tagger.

As was to be expected, a large effect of sentence boundary on syllable and vowel duration was observed. All examined syllables, including the first syllables of the prefixes [ʔun.tə], and [ʔy:.bə], were considerably lengthened when appearing in sentence-final, separated prefixes. The results confirm findings by Kohler (1983) and Silverman (1990), according to which sentence-final lengthening extends beyond the final syllable. Effects of sentence boundary were also found for prominence and spectral similarity, although not all syllables were affected equally. The interpretation of possible word boundary effects is not straightforward. A longer duration of free words might be expected due to effects of word-final lengthening (e.g., Beckman and Edwards, 1990) or polysyllabic shortening (e.g., Turk and Shattuck-Hufnagel, 2000; White, 2002), as bound prefixes were not followed by a word boundary and therefore appeared in longer words than the corresponding prepositions or conjunctions. Also, bisyllabic items had lexical stress on the first syllable as free words, but not as bound prefixes. On the other hand, there might have been counteracting influences of word frequency and accentual lengthening, as the verbs used were generally less frequent than the matching function words and tended to attract sentence focus. In our study, syllables in bound prefixes tended to be slightly shorter than when they occurred in segmentally identical prepositions or conjunctions, with the first syllable of the bisyllabic [ʔun.tə] and [ʔy:.bə] being affected more strongly than the second syllable. No influence was found for prominence and similarity values, and the word boundary effect was not significant in a separate investigation of the individual target syllables.



## EXPERIMENT 2: LEXICAL CLASS AND WORD FREQUENCY

Of the 2160 items recorded (8 words  $\times$  3 contexts  $\times$  3 lexical classes  $\times$  30 participants), 310 had to be omitted from the analysis. Results are based on the remaining 1850 items, which were investigated with regards to the factors lexical class (“dp,” “rp,” “da”) and word identity (“der masc.,” “der fem.,” “die sg.,” “die pl.,” “das,” “dem masc.,” “dem neut.,” “den”).

### Duration

In terms of syllable and vowel duration, demonstrative pronouns tended to be slightly longer than segmentally identical definite articles, with relative pronouns usually falling somewhere in between. This trend was especially noticeable for feminine “der” as well as masculine and neuter “dem.” Differences for “den,” masculine “der,” and singular “die” were less pronounced, while hardly any changes were observed for “das” and plural “die” (see **Table 6** for mean values). Two-Way ANOVAs were computed to examine the influence of word identity and lexical class on log-transformed syllable and vowel duration values. Significant effects ( $p < 0.0001$ ) were found for word identity [syllable duration:  $F_{(7, 1824)} = 147.8$ , vowel duration:  $F_{(7, 1824)} = 61.2$ ], lexical class [syllable duration:  $F_{(2, 1824)} = 123.2$ , vowel duration:  $F_{(2, 1824)} = 109.3$ ], and their interaction [syllable duration:  $F_{(14, 1824)} = 8.6$ , vowel duration:  $F_{(14, 1824)} = 11.0$ ]. Tukey’s HSD tests were used to further investigate the data. In terms of syllable as well as vowel duration, significant differences ( $p < 0.001$ ) were found between masculine and neuter “dem” and between masculine and feminine “der,” but not between singular and plural “die.” Significance

levels for the interaction between lexical class and word identity are given in **Table 7**.

### Prominence

Across items, prominences were higher for demonstrative pronouns than for relative pronouns and definite articles. Definite articles were minimally less prominent than relative pronouns. **Figure 5** shows results by lexical class for the individual words. Combinations of word identity and lexical class were analyzed using Wilcoxon rank sum tests (corrected for 276 comparisons, see **Table 8**). No significant differences between lexical classes were found for neuter “dem” or “den.” For all other items except masculine “dem,” demonstrative pronouns tended to receive higher prominence values than relative pronouns. Demonstrative pronouns were more prominent than definite articles for masculine and feminine “der” and masculine “dem.” While definite articles tended to be more prominent than relative pronouns for masculine “der,” singular and plural “die,” and “das,” an opposite trend was visible for masculine “dem.”

### Spectral similarity

Similarity levels were computed for segmentally identical items belonging to the same lexical class and produced by the same speaker in different contexts. Across words, definite articles (mean value: 0.814) appeared to be minimally less consistent in their pronunciation than demonstrative or relative pronouns (mean values: 0.823, 0.823). The difference, however, was only significant in Wilcoxon rank sum tests ( $W > 739,000$ ,  $p < 0.05$ , corrected

**Table 6 | Mean duration.**

Lexical class	der (masc.)	der (fem.)	die (sg.)	die (pl.)	das	dem (masc.)	dem (neut.)	den	All
dp	150.0	193.4	124.7	126.7	186.7	223.5	185.5	200.8	175.0
	96.9	139.4	79.4	71.6	72.7	84.2	85.6	82.7	89.8
rp	124.6	144.5	126.8	117.1	186.8	212.9	172.2	189.1	158.0
	78.3	101.8	68.0	64.4	60.5	76.7	76.5	73.9	74.8
da	118.2	104.3	100.0	112.3	182.0	161.9	141.9	150.7	133.6
	74.7	81.9	65.5	66.7	71.6	45.8	61.6	58.6	65.8

Mean duration values of demonstrative pronouns (dp), relative pronouns (rp) and definite articles (da) in milliseconds for syllables (above) and vowels (below).

**Table 7 | Interaction of lexical class and word identity.**

Lexical class	der (masc.)	der (fem.)	die (sg.)	die (pl.)	das	dem (masc.)	dem (neut.)	den
dp vs. rp	$p < 0.05$	$p < 0.0001$	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
	n.s.	$p < 0.0001$	n.s.	n.s.	$p < 0.05$	n.s.	n.s.	n.s.
dp vs. da	$p < 0.01$	$p < 0.0001$	$p < 0.001$	n.s.	n.s.	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$
	$p < 0.01$	$p < 0.0001$	n.s.	n.s.	n.s.	$p < 0.0001$	$p < 0.0001$	$p < 0.0001$
rp vs. da	n.s.	$p < 0.0001$	$p < 0.0001$	n.s.	n.s.	$p < 0.0001$	$p < 0.01$	$p < 0.01$
	n.s.	n.s.	n.s.	n.s.	n.s.	$p < 0.0001$	$p < 0.01$	$p < 0.05$

Adjusted  $p$ -values (Tukey HSD) for comparisons of syllable duration (above) and vowel duration (below) between demonstrative pronouns (dp), relative pronouns (rp) and definite articles (da).



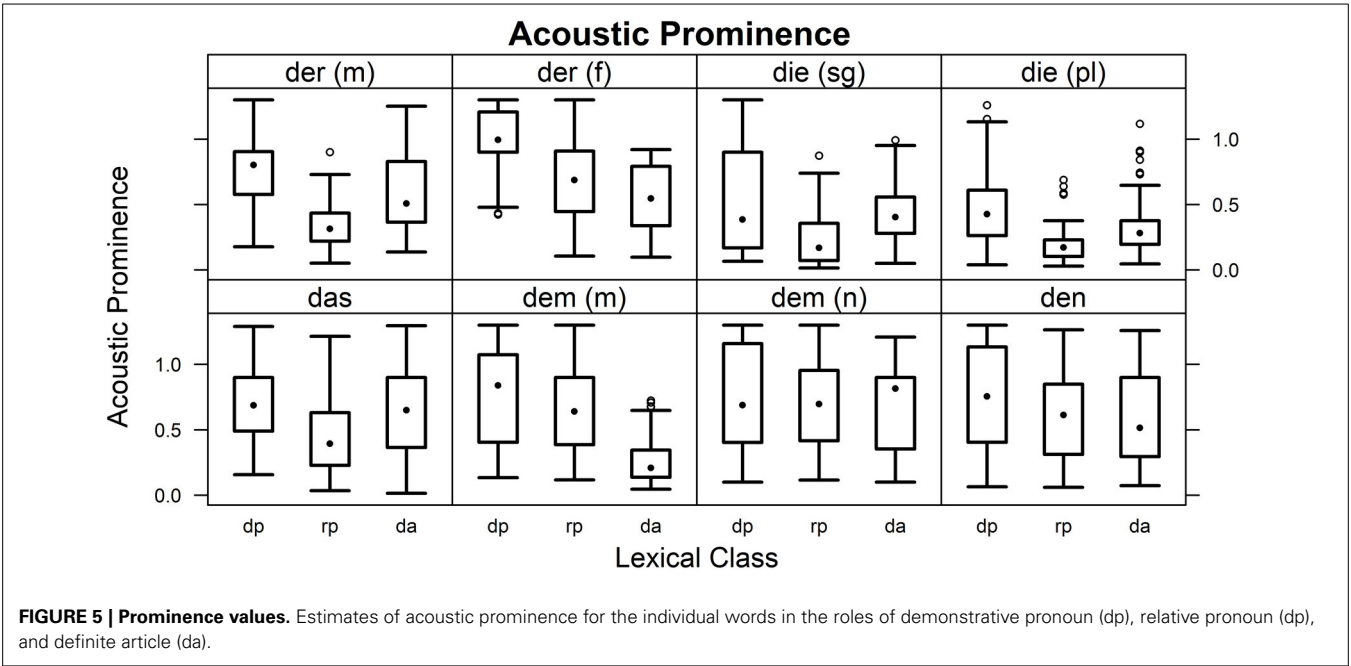


Table 8 | Prominence statistics for differences in lexical class.

Lexical classes	der (masc)	der (fem)	die (sing)	die (plur)	das	dem (masc)	dem (neut)	den
dp vs. rp	4922.5****	5067.0****	3935.5***	4934.5****	4840.5****	3089.0 (n.s.)	3286.5 (n.s.)	3061.5 (n.s.)
dp vs. da	3851.0**	5689.5****	3099.5 (n.s.)	4022.5 (n.s.)	3353.5 (n.s.)	5054.5****	3388.0 (n.s.)	3687.0 (n.s.)
da vs. rp	4937.5****	2135.5 (n.s.)	4739.0 ****	5144.4 ****	4566.0 **	793.0 ****	2887.5 (n.s.)	2567.5 (n.s.)

W values with significance levels for comparisons between definite pronouns (dp), relative pronouns (rp) and definite articles (da), n.s.:  $p \geq 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

for 3 comparisons) when similarities were calculated regardless of gender or class. No effects were found when word identity as well as segmental identity was controlled (corrected for 3 comparisons), or when lexical classes were compared separately for individual word identities (corrected for 276 comparisons). We also examined similarity levels between words belonging to different lexical classes (paired for speaker, word identity, and context). Here, we found a significant difference between similarity measures of relative and demonstrative pronouns on the one hand and relative pronouns and definite articles on the other (mean values: 0.861 vs. 0.850,  $W = 155,459.5$ ,  $p < 0.05$ , corrected for 3 comparisons). In separate comparisons for individual word identities (corrected for 276 comparisons), this tendency was only confirmed for masculine “dem” (mean values: 0.893 vs. 0.831,  $W = 2651$ ,  $p < 0.001$ ).

Discussion

Definite articles were expected to have smaller duration values than segmentally identical relative or demonstrative pronouns due to effects of frequency and predictability. Not only are they much more common than the other lexical classes, the carrier sentences for the pronouns were specifically constructed to mirror the phonetic context of definite articles found in other sentences, probably increasing their artificiality and reducing

the predictability of the target words. According to exemplar-theoretic approaches, definite articles might also be more strongly adapted to their surroundings, which would lead to lowered spectral similarity values across contexts. However, differences in pronunciation cannot always be explained by lemma frequency, and lexical classes may vary in the degree to which they can be emphasized. For instance, Jurafsky et al. (2000) found that although the English word “that” was most commonly used as a demonstrative pronoun, it tended to be longer in this function than when it was produced as a segmentally identical relative pronoun, complement, or determiner. In order to monitor for differences in emphasis, we also analyzed the target words’ level of acoustic prominence in relation to their immediate context. In our investigation, we discovered significant differences between all three lexical classes in terms of syllable and vowel duration. Although these differences were not contradictory to lemma frequency effects, they did not mirror the fact that in German, frequency differences between the two types of pronouns are minimal compared to their difference to definite articles. The comparatively high duration of demonstrative pronouns was probably due to their semantic role, as it is their function to point out and emphasize the entity to which they refer. Results for acoustic prominence confirm that participants tended to emphasize demonstrative pronouns more strongly than relative pronouns

or definite articles. Contrary to our expectations, we found only minimal effects and no consistent patterns in terms of spectral similarity within and between lexical classes.

A closer examination of the data revealed that the individual target words varied in the ways and extent to which they were affected by changes in lexical class. Duration differences were most stable in comparisons between demonstrative pronouns and definite articles. Relative pronouns often tended to be closer in duration to demonstrative pronouns than to definite articles. Plural “die” showed no duration effects whatsoever, and the only significant duration effect found for “das” was a slight difference in vowel duration between relative and demonstrative pronouns. Concerning acoustic prominence, it was striking that while any significant differences between demonstrative pronouns and definite articles was accompanied by significant effects of syllable and vowel duration, several words showed prominence differences between relative pronouns and the other two categories without any corresponding duration effects. Although relative pronouns were generally longer than definite articles, prominence levels tended to be lower, with only masculine “dem” showing a significant effect in the opposite direction. Only singular “die” showed contradictory duration and prominence results which were both significant. The conflicting prominence findings may have resulted from the difficulty in controlling the context of the target items. As relative pronouns are generally used to introduce relative clauses, the syllables preceding them tended to be clause-final and therefore subject to final lengthening. It is very likely that relative pronouns received particularly low prominence ratings by the tagger due to their relatively prominent preceding context. In the case of feminine “der,” masculine and neuter “dem,” and one sentence used for “den,” possible context lengthening was avoided by placing the relative pronouns in prepositional phrases. For these words, there was indeed no tendency for relative pronouns to be less prominent than definite articles, and prominence differences were supported by differences in syllable and vowel duration.

## SUMMARY

This paper describes results from two experiments designed to disentangle various influences on syllable pronunciation in German. In the first experiment, we found clear differences due to word stress and sentence boundaries, while effects of sentence stress and word boundaries were smaller in size and less consistent across stimuli. In the second experiment, differences between segmentally identical demonstrative pronouns, relative pronouns, and definite articles were found that could be related to lemma frequency, semantic function, and sentence structure. In both experiments, duration was shown to be the most robust of the investigated cues for disambiguating word meanings. Measures of acoustic prominence added valuable information on how strongly syllables were emphasized, but also proved to be highly sensitive to differences in context. Finally, an examination of spectral similarity revealed that syllables in lexically stressed prefixes were less variable across contexts and closer in pronunciation to sentence-final realizations than unstressed prefixes. Separate investigations of individual target syllables often failed to reach significance in terms of acoustic prominence and spectral similarity, suggesting

that other influences may also have been of importance. Especially prominence and similarity measures often failed to reach significance in these detailed analyses. A larger study covering a greater number of contexts and using a separate quasi-random order of sentences for each speaker, possibly followed by a perception study to confirm the results, might lead to more robust findings.

## ACKNOWLEDGMENTS

This study was funded by the German Research Foundation (DFG), Priority Program 1234, grant MO 597/4. We would like to thank Natalie Lewandowski for providing the Matlab scripts used for calculating spectral similarity measures. We also thank the two anonymous reviewers for their insightful and constructive comments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00500/abstract>

## REFERENCES

- Aylett, M., and Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang. Speech* 47, 31–56. doi: 10.1177/00238309040470010201
- Baker, R. E., and Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Lang. Speech* 52, 391–413. doi: 10.1177/0023830909336575
- Baroni, M., and Kilgariff, A. (2006). “Large linguistically-processed web corpora for multiple languages,” in *Proceedings of EAACL 2006*. (Trento).
- Beckman, M. E., and Edwards, J. (1990). “Lengthenings and shortenings and the nature of prosodic constituency,” in *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech*, eds J. Kingston and M. E. Beckman (Cambridge: Cambridge University Press), 152–178.
- Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *J. Mem. Lang.* 60, 92–111. doi: 10.1016/j.jml.2008.06.003
- Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *J. Acoust. Soc. Am.* 113, 1001–1024. doi: 10.1121/1.1534836
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott Int.* 5, 341–345.
- Cambier-Langeveld, T., and Turk, A. E. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *J. Phon.* 27, 255–280. doi: 10.1006/jpho.1999.0096
- Campbell, N., and Beckman, M. (1997). “Stress, prominence, and spectral tilt,” in *Proceedings of ESCA Workshop “Intonation: Theory, Models and Applications” 1997*. (Athens).
- Cho, T., and Keating, P. (2009). Effects of initial position versus prominence in English. *J. Phon.* 37, 466–485. doi: 10.1016/j.wocn.2009.08.001
- Claßen, K., Dogil, G., Jessen, M., Marasek, K., and Wokurek, W. (1998). Stimmqualität und Wortbetonung im Deutschen. *Linguist. Ber.* 174, 202–245.
- Coyne, B., and Sproat, R. (2001). “WordsEye: an automatic text-to-scene conversion system,” in *Proceedings of SIGGRAPH 2001* (Los Angeles, CA).
- Dogil, G., and Williams, B. (1999). “The phonetic manifestation of word stress,” in *Word Prosodic Systems in the Languages of Europe*, ed H. van der Hulst (Berlin; New York, NY: Mouton de Gruyter), 273–334.
- Jurafsky, D., Bell, A., and Girand, C. (2000). “The role of the lemma in form variation,” in *Laboratory Phonology 7*, eds C. Gussenhoven and N. Warner (Berlin; New York, NY: Mouton de Gruyter), 3–34.
- Kleber, F., and Klippahn, N. (2006). An acoustic investigation of secondary stress in German. *Arbeitsberichte Institut für Phonetik Kiel* 37, 1–18.
- Kohler, K. J. (1983). Prosodic boundary signals in German. *Phonetica* 40, 89–134. doi: 10.1159/000261685

- Kohler, K. J. (1987). "The linguistic functions of F0 peaks," in *Proceedings of ICPHS 1987*. (Tallinn).
- Lewandowski, N. (2011). *Talent in Nonnative Phonetic Convergence*. Ph.D. dissertation, University of Stuttgart, Stuttgart.
- Lintfert, B. (2010). *Phonetic and Phonological Development of Stress in German*. Ph.D. dissertation, University of Stuttgart, Stuttgart.
- Mooshammer, C., Fuchs, S., and Fischer, D. (1999). "Effects of stress and tenseness on the production of CVC syllables in German," in *Proceedings of ICPHS 1999*. (San Francisco, CA).
- Okobi, A. O. (2006). *Acoustic Correlates of Word Stress in American English*. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge.
- Plag, I., Kunter, G., and Schramm, M. (2011). Acoustic correlates of primary and secondary stress in North American English. *J. Phon.* 39, 362–374. doi: 10.1016/j.wocn.2011.03.004
- Pluymaekers, M., Ernestus, M., and Baayen, H. R. (2005). Lexical frequency and acoustic recution in spoken Dutch. *J. Acoust. Soc. Am.* 118, 2561–2569. doi: 10.1121/1.2011150
- R Development Core Team. (2010). *R: A Language and Environment for Statistical Computing*. Available online at: <http://www.r-project.org>
- Samlowski, B., Wagner, P., and Möbius, B. (2012). "Disentangling lexical, morphological, syntactic and semantic influences on German prosody - Evidence from a production study," in *Proceedings of Interspeech 2012*. (Portland, OR).
- Samlowski, B., Wagner, P., and Möbius, B. (2013). "Effects of lexical class and lemma frequency on German homographs," in *Proceedings of Interspeech 2013*. (Lyon).
- Schneider, K., and Möbius, B. (2007). "Word stress correlates in spontaneous child-directed speech in German," in *Proceedings of Interspeech 2007*. (Antwerp).
- Silverman, K. E. A. (1990). "The separation of prosodies: comments on Kohler's paper," in *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech*, eds J. Kingston and M. E. Beckman (Cambridge: Cambridge University Press), 139–151.
- Sluijter, A. M. C., and van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *J. Acoust. Soc. Am.* 100, 2471–2485. doi: 10.1121/1.417955
- Tamburini, F., and Wagner, P. (2007). "On automatic prominence detection for German," in *Proceedings of Interspeech 2007*. (Antwerp).
- Turk, A. E., and Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *J. Phon.* 28, 397–440. doi: 10.1006/jpho.2000.0123
- Wade, T., and Möbius, B. (2007). "Speaking rate effects in a landmark-based phonetic exemplar model," in *Proceedings of Interspeech 2007*. (Antwerp).
- Wagner, P. (2002). *Vorhersage und Wahrnehmung deutscher Betonungsmuster*. Ph.D. dissertation, Bonn University, Bonn.
- White, L. (2002). *English Speech Timing: A Domain and Locus Approach*. Ph.D. dissertation, University of Edinburgh, Edinburgh.
- Widera, C., Portele, T., and Wolters, M. (1997). "Prediction of word prominence," in *Proceedings of Eurospeech 1997*. (Rhodes).
- Windmann, A., Jauk, I., Tamburini, F., and Wagner, P. (2011). "Prominence-based prosody prediction for unit selection speech synthesis," in *Proceedings of Interspeech 2011*. (Florence).

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 January 2014; accepted: 07 May 2014; published online: 27 May 2014.

Citation: Samlowski B, Möbius B and Wagner P (2014) Phonetic detail in German syllable pronunciation: influences of prosody and grammar. *Front. Psychol.* 5:500. doi: 10.3389/fpsyg.2014.00500

This article was submitted to *Language Sciences*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Samlowski, Möbius and Wagner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Use what you can: storage, abstraction processes, and perceptual adjustments help listeners recognize reduced forms

Katja Poellmann<sup>1,2\*</sup>, Holger Mitterer<sup>1†</sup> and James M. McQueen<sup>1,3</sup>

<sup>1</sup> Language Comprehension Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

<sup>2</sup> International Max Planck Research School for Language Sciences, Nijmegen, Netherlands

<sup>3</sup> Behavioural Science Institute and Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands

## Edited by:

Ulrike Domahs, University of Marburg, Germany

## Reviewed by:

LouAnn Gerken, University of Arizona, USA

Judith Koehne, Saarland University, Germany

## \*Correspondence:

Katja Poellmann, Department of Speech-Language Pathology and Audiology, Northeastern University, 360 Huntington Avenue, 226 FR, Boston, MA 02115, USA  
e-mail: k.poellmann@neu.edu

## † Present address:

Katja Poellmann, Department of Speech-Language Pathology and Audiology, Northeastern University, Boston, USA;  
Holger Mitterer, Department of Cognitive Science, University of Malta, Msida, Malta

Three eye-tracking experiments tested whether native listeners recognized reduced Dutch words better after having heard the same reduced words, or different reduced words of the same reduction type and whether familiarization with one reduction type helps listeners to deal with another reduction type. In the exposure phase, a segmental reduction group was exposed to /b/-reductions (e.g., *minderij* instead of *binderij*, “book binder”) and a syllabic reduction group was exposed to full-vowel deletions (e.g., *p’raat* instead of *paraat*, “ready”), while a control group did not hear any reductions. In the test phase, all three groups heard the same speaker producing reduced-/b/ and deleted-vowel words that were either repeated (Experiments 1 and 2) or new (Experiment 3), but that now appeared as targets in semantically neutral sentences. Word-specific learning effects were found for vowel-deletions but not for /b/-reductions. Generalization of learning to new words of the same reduction type occurred only if the exposure words showed a phonologically consistent reduction pattern (/b/-reductions). In contrast, generalization of learning to words of another reduction type occurred only if the exposure words showed a phonologically inconsistent reduction pattern (the vowel deletions; learning about them generalized to recognition of the /b/-reductions). In order to deal with reductions, listeners thus use various means. They store reduced variants (e.g., for the inconsistent vowel-deleted words) and they abstract over incoming information to build up and apply mapping rules (e.g., for the consistent /b/-reductions). Experience with inconsistent pronunciations leads to greater perceptual flexibility in dealing with other forms of reduction uttered by the same speaker than experience with consistent pronunciations.

**Keywords:** reduction, word-specificity, generalization, learning, adaptation, eye-tracking

## INTRODUCTION

In casual speech, speakers tend to articulate in a sloppy way. They frequently reduce words by slurring and even omitting segments or syllables (Ernestus, 2000; Patterson et al., 2003; Johnson, 2004; Mitterer and McQueen, 2009). A given native Dutch speaker may for example reduce the /b/ in *bandiet* “bandit” to [m] or leave out the first vowel in *kanaal* “canal” (Schuppler et al., 2011). Listeners might get used to such pronunciation habits; they may recognize a reduced word better the second time and they may be able to adjust rapidly to new forms of reduction produced by the same speaker. The present study investigates *whether* listeners adapt to a given reduction type (/b/-reductions or full-vowel-deletions) and, if so, *how* they adapt by asking if they can apply their knowledge to previously unheard reduced words of the same reduction type and/or of the other reduction type. Put another way, the present study tests word-specific learning effects as well as generalization of learning within and across reduction types.

Listeners are usually not aware that they encounter numerous reduced word forms every day (Kemps et al., 2004; Ernestus

and Warner, 2011). They use the information provided by the sentence context or also the wider discourse context to predict and, if necessary, restore the upcoming word (Ernestus et al., 2002; Brouwer et al., 2013). On a lower level, listeners are also able to exploit the fine phonetic detail present in reduced forms to distinguish for instance between a reduced form [spɔ:t] of *support* and the unreduced form [spɔ:t] *sport* (Manuel, 1992).

Another mechanism which listeners may use to recognize reduced forms better is adaptation, as perceptual learning may be especially important when the conditions for spoken-word recognition become challenging.

Adaptation, for instance, has been found to play a crucial role in recognizing regional and foreign-accented speech (Clarke and Garrett, 2004; Floccia et al., 2006; Mitterer and McQueen, 2009). Listeners are able to adapt rapidly to these deviant pronunciations and can apply their acquired knowledge to the way they process other words (Witteman et al., 2013).

The present study tests whether a similar adaptation process also takes place when listeners encounter reduced words in

their native language. Like regional and foreign-accented words, reduced words are also variants of canonical pronunciations, but the reduction types chosen for investigation in the present study (/b/-reductions and full-vowel-deletions) were not regionally marked. In contrast to regional and foreign accents, reductions affect predominantly unstressed segments and syllables. They are therefore probably less salient. This might make it harder for listeners to adapt to reduced speech than to regional or foreign-accented speech.

The present study investigates potential adaptation processes and their possible constraints. Consider a Dutch listener hearing the word *paraat* “ready” pronounced as *p’raat*. Different patterns of adaptation are possible that vary in how general they are. First, no adaptation whatsoever may be found. Second, the listener may find it easier to recognize a second instance of the same word with the same reduction pattern. This would be similar to the recognition benefits for words repeated in the same voice that provide some of the evidence for episodic models of word recognition (Nygaard et al., 1994; Goldinger, 1996, 1998; Nygaard and Pisoni, 1998). Third, listeners may learn that this speaker generally deletes vowels in unstressed syllables. This abstractionist learning may be quite specific, so that only very similar reductions to *p’raat* benefit (e.g., *Parijs* “Paris” produced as *P’rijs*; note that the Dutch rendition is stressed on the second syllable) or it may include reductions of unstressed vowels in other contexts (e.g., *kanaal* “canal” produced as *k’naal*). The strongest possible generalization would be that the listener assumes that this speaker reduces a great deal and hence finds it easier to recognize any kind of reduction uttered by the speaker.

Finding a word-specific learning effect, that is, better recognition of a reduced word on hearing it for the second time compared to the first time, would be evidence for episodic storage of reduced forms. In contrast, observing generalization of learning to new words of the same reduction type (e.g., generalization from *p’raat* to *P’rijs* or *k’naal*) would indicate that an abstraction process is taking place and that it occurs at a prelexical level. Storing reduced forms alone cannot account for easier recognition of previously unheard reduced words (McQueen et al., 2006; Cutler et al., 2010). In a purely episodic account of lexical access, there is no way to adjust weights of sublexical units like segments and syllables to build up rules that capture regular reduction processes (e.g., “Potentially restore a bilabial nasal in an unstressed syllable to a bilabial voiced stop if followed closely by another nasal”). Finding generalization of learning to new words of the same reduction type would thus support the claim that there is abstraction in lexical access. Observing generalization of learning from one reduction type to another may also be evidence for abstraction—if there is enough similarity between the reduction types to abstract over the respective mapping rules. Consider, for example, two types of prefix reductions, such as *ge-/gə/* → */g/* and *be-/bə/* → */b/* in German. An abstraction rule may be: “Potentially insert a schwa after an initial voiced stop” (instead of “...after an initial voiced velar/bilabial stop”). However, should generalization of learning across reduction types be found for very different reduction types, such as the /b/-reductions and full-vowel-deletions examined here, this would more likely indicate a non-specific adjustment and be evidence for the flexibility of the

perceptual system. That is, instead of specific adaptation processes (storage of reduced forms and/or abstraction of reduction rules), listeners could make a more general adjustment to the current talker’s speaking style.

To test these possible adaptation effects, the printed-word eye-tracking paradigm (McQueen and Viebahn, 2007) was used. In the exposure phase, one group of participants was exposed to segmental reductions, another group was exposed to syllabic reductions and a third group was exposed only to canonical pronunciations. The first group, the segmental reduction group, heard /b/-reductions, where the word-initial /b/ was reduced to a bilabial nasal (e.g., *minderij* instead of *binderij* “book binder”). The second group, the syllabic reduction group, heard words in which the first, unstressed full vowel was deleted (e.g., *p’raat* instead of *paraat* “ready”). The third group, the control group, heard the same words as the two experimental groups during the exposure phase but all in unreduced form (e.g., *binderij* and *paraat*).

In order to assess the frequency with which our chosen reduction types (/b/-reductions and full-vowel-deletions) occur in spontaneous speech, we conducted a corpus study following the principles of Pluymaekers et al. (2005). First, all sound files containing a /b/-initial word with a nasal in third position and an unstressed first syllable were extracted from the Corpus of Spoken Dutch (Oostdijk, 2000). Per word type (this notion here not only describes words belonging to different lemmas but also different word forms of one lemma, e.g., an inflected verb form or the plural of a noun) only one token was randomly chosen to determine its phonetic realization. Out of 65 word types, six showed a /b/ → [m] reduction in the first segment (i.e., 9.2% of the considered cases). A similar analysis was conducted to assess the frequency of full-vowel-deletions in initially unstressed words. The vowel was deleted in eight out of 66 word types (i.e., in 12.1%) containing either a voiceless plosive (/p/, /k/) or a voiceless velar fricative (/x/) in first position and an alveolar nasal or liquid in third position. This was also the segmental structure used in the syllabic reduction condition. The chosen reduction types were thus indeed real-world phenomena and comparable in terms of frequency.

These two reduction types were chosen to examine adaptation to two different-sized linguistic units, the phoneme and the syllable, and the possible interaction of the adaptation effects. An earlier study showed that listeners adapt to syllabic reductions involving a morpheme: After exposure to words containing the reduced prefix *ver-* (realized as [f:]), Dutch listeners recognized previously unheard reduced *ver*-words better than a control group (Poellmann et al., under revision). In the present study, we test whether this is also the case for non-morphemic syllables. The deletion of the unstressed, full vowel in CVC-initial words like *paraat* always leads to a reduction in the number of syllables, which is why this reduction type is called “syllabic.” A pure comparison of morphemic and non-morphemic reductions, however, turned out to be impossible in Dutch. Ideally, one would like to compare a morphemic reduction type (that only affects one specific morpheme, i.e., the same strings of segments, such as Dutch *ge-*) to a non-morphemic reduction type that also only affects one specific string of segments (e.g., *pa-*). The Dutch



lexicon, however, does not contain enough words starting with one specific unstressed non-morphemic syllable to conduct such an experiment. This constraint on the (non-)morphemic status hence leads inevitably to higher variability in the segmental structure of the CVC-targets compared to the *ver*-targets examined in Poellmann et al. (under revision). This difference in the degree of consistency with which words are reduced in the two conditions allowed us to ask whether phonological consistency determines which adaptation processes (e.g., storage, abstraction rules, general flexibility) listeners are able to use.

In the test phase, all three groups of participants heard /b/-reductions and vowel-deletions. The reduced words were either the same as in the exposure phase (in Experiments 1 and 2) or different (in Experiment 3). If listeners adapt to a given reduction type and if they can transfer this knowledge to new words (Experiment 3) and/or to other reduction types (Experiments 1–3), participants in the experimental groups should recognize reduced words better than participants in the control group.

Regardless of the specifics concerning the reduction (such as size of the reduced unit or input consistency), it seems plausible that a reduced word can be recognized more easily if it is encountered a second time. We therefore expect to find word-specific learning effects for both /b/-reductions and vowel-deletions.

Moreover, we predict that learning about /b/-reductions generalizes to new words that are reduced in the same way. Such generalization effects have been observed for a similar kind of /b/-reduction where the word-initial voiced stop was reduced to a labio-dental approximant [v] (Poellmann et al., under revision) and for learning about segmental idiosyncrasies (McQueen et al., 2006). In the McQueen et al. (2006) study, listeners adapted to an ambiguous sound (between /s/ and /f/) and transferred their knowledge to previously unheard minimal pairs that only differed in containing either /s/ or /f/.

The predictions concerning within-reduction-type generalizations for full-vowel-deletions are less clear. The constraint on the (non-)morphemic status of the syllable leads to higher variability in the segmental structure of the CVC-targets compared to the /b/-targets. If the input has to be highly consistent for the creation of abstract mapping rules, we might not observe generalization of learning.

The two reduction types under investigation differ in several respects, such as the degree of reduction (weakening of the [b] vs. deletion of the vowel), in the segment that is reduced (bilabial voiced stop vs. full vowel) and in the position the reduced segment occurs (first position for /b/-reductions vs. second position for vowel-deletions). In order to observe generalization of learning across reduction types, listeners would hence have to adapt on a fairly global level. However, such global adjustments to challenging listening conditions have been observed before (Brouwer et al., 2012; McQueen and Huettig, 2012).

## EXPERIMENT 1

The aim of Experiment 1 was to test whether listeners are able to recognize segmental and syllabic reductions better when they have already encountered the same words in reduced form before. Experiment 1 also asked whether learning about reductions might generalize from one reduction type to another (i.e.,

from /b/-reductions to full-vowel deletions and/or vice versa). In the exposure phase, one group was exposed to /b/-reductions (segmental reduction group), a second group was exposed to full-vowel deletions (syllabic reduction group), while a third group was exposed to canonical forms only (control group). In the test phase, all three groups were tested on reduced-/b/ words and vowel-deleted words. Importantly, these reduced words had already occurred in reduced or canonical form (depending on the group) in the exposure phase. If listeners can adapt to reduced words, the segmental reduction group should recognize the reduced-/b/ words better than the syllabic reduction group and the control group because of their previous exposure to these words in reduced form. The same holds for participants in the syllabic reduction group: If they can adapt to vowel-deleted words, they should perform better on these words than the segmental reduction group and the control group. If listeners can additionally transfer their knowledge about one reduction type to another, the segmental reduction group should outperform the control group on the vowel-deleted words and the syllabic reduction group should outperform the control group on the reduced-/b/ words.

## METHODS

### Participants

Seventy-five participants of the Max Planck Institute's subject pool, all native speakers of Dutch, were paid to take part. All reported normal hearing and normal or corrected-to-normal vision.

### Design

Participants were randomly assigned to one of three groups: a segmental reduction group, a syllabic reduction group and a control group. They listened to sentences, saw four printed words on a computer screen and were asked to click on the word that occurred in the sentence. Improved word recognition in a visual-world eye-tracking experiment can be reflected by faster and more accurate mouse clicks on the target word as well as higher fixation proportions toward the target and away from the similar sounding competitor. We thus measured Reaction Times (RTs) and accuracy of mouse clicks and fixation behavior.

In the exposure phase, participants were exposed to words that were potentially reduced (see the experimental exposure trials in Table 1) but which did not appear on the screen. Instead, they saw (and had to click on) target words that occurred later in the sentences. All three groups were also exposed to unreduced /m/- and unreduced consonant-cluster-words (e.g., /matros/ *matroos* “sailor” and /knɔflok/ *knoflook* “garlic”); they also had to click on these filler stimuli.

In the test phase, all three groups heard reduced /b/-words and vowel-deleted words in the experimental trials. These were the same words as had appeared in the exposure phase (e.g., [mindərəi] instead of [bindərəi] *binderij* “book binder” and [prat] instead of [parat] *paraat* “ready”). All groups also heard new canonical /m/- and new canonical consonant-cluster words. The reduced /b/-words, the vowel-deleted words, the unreduced /m/-word fillers and the consonant-cluster filler words were all

**Table 1 | Experimental design and types of stimuli in Experiments 1 and 2.**

	Trial type	Canonical word-form	Segmental reduction group	Syllabic reduction group	Control group
			/b/ → [m]	Full vowel deletion	No reduction
Exposure phase	Experimental	/bɪndərəɪ/	<b>[mɪndərəɪ]</b>	[bɪndərəɪ]	[bɪndərəɪ]
	Filler	/matros/	[matros]	[matros]	[matros]
	Experimental	/parat/	[parat]	<b>[prat]</b>	[parat]
	Filler	/knɔflok/	[knɔflok]	[knɔflok]	[knɔflok]
Test phase	Experimental	/bɪndərəɪ/		<b>[mɪndərəɪ]</b>	
	Filler	/muras/		[muras]	
	Experimental	/parat/		<b>[prat]</b>	
	Filler	/xlans/		[xlans]	

Reduced segments are marked bold. The potentially reduced /b/-initial words and the vowel-deleted words of the exposure phase were repeated in reduced form in the test phase.

targets and were therefore displayed on the computer screen in (canonical) orthographic form.

### Materials

The target words (i.e., the words participants had to click on) appeared toward the end of spoken sentences. Each target word occurred in a different sentence context not containing any further /b/s in unstressed syllables or any further unstressed CVC-sequences which would result in legal consonant clusters when omitting the vowel. The potentially reduced item occurred before the target word in the experimental trials (e.g., *Pas in een [b]/[m]inderij wordt een boek of tijdschrift afgemaakt* “Only at a book binder, a book or **magazine** gets finished,” where bold font indicates the target word and underlining marks the potentially reduced critical item). This was done to prevent participants from clicking on the same words twice, once in the exposure phase and once in the test phase. In the test phase, the semantic contexts preceding the target words were kept uninformative (e.g., *Het tekstverwerkingprogramma kende het woordje [m]inderij niet* “The word processor did not know the word **book binder**”). During each sentence, there were always four printed words on the screen. In the test trials, these were a /b/-word, a /m/-word, a CVC-word and a consonant-cluster word (see **Figure 1** for an example display).

The test phase consisted of 48 experimental trials containing either /b/-targets or CVC-targets and 48 filler trials containing either /m/-targets or CC-targets. For each type of target word (/b/-target, /m/-target, CVC-target, and consonant-cluster-target), 24 target-competitor pairs were selected (see Table S1 for the /b/-targets, the CVC-targets, and their respective competitors). If a /b/-word was the target, a /m/-word was the competitor and vice versa. The same holds for CVC- and consonant-cluster-targets. All /b/- and /m/-initial words contained an unstressed first syllable. In second position, any vowel including schwa could occur followed by a nasal in third position. The latter condition was necessary for all /b/-targets to motivate nasalization at the beginning of the word. However, there are not sufficient /m/-initial words in Dutch containing a nasal in third position to create perfectly matched pairs of /b/-targets and /m/-competitors. Ideally, /b/-words and /m/-words should be as similar as possible

with as much overlap in the reduced forms as possible (e.g., *binderij* “book binder” pronounced as [mɪndərəɪ] overlaps in the first two syllables with [mɪndərjarəx] *minderjarig* “under-age”). Due to the infrequent occurrence of a nasal in third position following an /m/ in first position, the /m/-targets contained a random consonant in third position (and so did the corresponding /b/-competitors; e.g., *moeras* “swamp” and *boerin* “farmer’s wife”). Target-competitor pairs were further matched in terms of number of syllables, stress pattern and word frequency [taken from SUBTLEX-NL (Keuleers et al., 2010)] as much as possible (see Table S1).

The principles of as much overlap and similarity as possible between targets and competitors also applied to the (reduced) CVC- and (unreduced) consonant-cluster-words. CVC-words started with an open syllable, consisting of a voiceless consonant (either /p/, /k/, or /x/) and a full vowel, followed by a liquid or /n/ in third position (e.g., *paraat* “ready”), so that the sequence resulting from vowel deletion would be a phonotactically legal consonant cluster in Dutch. The consonant-cluster words started with the same voiceless consonants directly followed by a liquid or [n] (e.g., *praat* “talk”). While the stress of the CVC-words was on the second syllable, the consonant-cluster-words were stressed on the first syllable, so that both word types were matched on stress pattern when the full vowel of the CVC-words was deleted (e.g., p’RAAT for paRAAT “ready” and PRAAT “talk”). Again, target-competitor pairs were matched on number of syllables (in the reduced form) and word frequency (see Table S1).

The exposure phase consisted of 96 trials in total. Half of them were filler trials containing /m/-targets or CC-targets. The 48 experimental trials contained potentially reduced /b/-words or CVC-words that did not appear on the screen. The only constraint for the target-“competitor” pairs on the screen was that they did not overlap.

### Stimulus construction

Digital recordings of the stimuli were made by a female native speaker of Dutch in a sound-proof booth, sampling at 44.1 kHz. She was instructed to produce the sentences in a casual way, not just reading them aloud. For sentences containing canonically pronounced /b/-targets, an additional set containing reduced

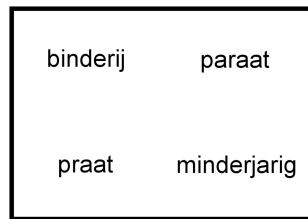


FIGURE 1 | Example display of a test trial in Experiments 1–3.

forms was created by replacing the /b/ with an /m/ from a word with the same vowel context. The spliced parts were adjusted in pitch (with PSOLA in PRAAT, Boersma and Weenink, 2010) and intensity to their new context. The transitions in amplitude preceding and following the spliced-in [m]s were smoothed where necessary in order to reduce splicing artifacts. The set of sentences containing reduced CVC-words was created by cutting out the first (unstressed) vowel of the recorded versions of these words with intact vowels. Sentence contexts were thus identical across the reduced and unreduced forms of each target word. Filler sentences containing /m/- and consonant-cluster-targets were not manipulated.

### Procedure

Participants were seated in a sound-attenuated booth at a comfortable viewing distance from the computer screen. Eye movements were monitored using an SR Research EyeLink 1000 set-up, sampling at 1 kHz. The auditory stimuli were presented to the participants over headphones. Prior to the experiment, participants received written instructions that informed them that they would see four printed words on the screen and asked them to click on the word that occurred in the sentence.

At the beginning of each trial, a fixation cross appeared in the center of the screen for 500 ms. Four printed words (in a 25-point Arial font) were then presented. After 1500 ms, the auditory stimulus was played. As soon as participants had listened to the entire sentence and had clicked with the mouse on the screen, the following trial was initiated. Every 10 trials, a drift correction was carried out. Participants had the opportunity to take a break after every 50th stimulus. The experiment started with six practice trials. The 96 exposure trials in random order were followed by 96 test trials in random order. Randomization was different for each participant. An experimental session took approximately 25 min.

## RESULTS

### Exclusion criteria

Mouse click responses (reaction time and accuracy data) and eye movements served as dependent variables. For the eye-tracking data, we analyzed the data from the participant's right eye. For the analysis of the eye-tracking data, a total of 2.9% of the trials were excluded, because participants either appeared to have looked away from the screen (2.0%) or failed to click on the target or the potentially confusable competitor (0.9%). Clicks on the competitor were not excluded from all of the analyses, as the competitors sometimes better fitted the exact auditory input with reduced forms than the targets. For instance, reduced *p'raat*

better fitted the canonical form of the competitor *praat* than the canonical form of the target *paraat*. Furthermore, the semantics of the test sentences did not make clear which word was the target. In the case of minimal pairs such as *paraat* and *praat*, participants thus never received disambiguating information about which of the two words they should click on. Therefore, clicks on competitors were not regarded as errors in the analyses of the eye-tracking and the reaction time data. Note also that excluding trials from the eye-tracking analysis in which participants clicked on the competitor would invalidate any learning effects. Presumably, participants look more at the competitor when they click on it. Excluding these trials would result in a greater preference for the target over the competitor and would thus misleadingly indicate a greater learning effect than was actually present. Moreover, the focus in the RT analyses is on the comparisons across the three exposure groups; these comparisons are thus orthogonal to any differences between targets and competitors. Click responses to competitors, however, were regarded as incorrect in the analysis of the accuracy scores.

The upper part of **Table 2** displays descriptive statistics on RTs for trials in which participants clicked either on the target or on the phonological competitor in the test phase of Experiment 1. Participants in the syllabic reduction group took longer to respond than participants in the segmental or no-reduction group. Participants, however, were not asked to respond as fast as possible. Some participants chose to do so; others waited for the sentence to finish before giving a response. The high standard deviation (SD) values reflect these different strategies. Extreme cases, that is, trials in which participants responded either too fast or too slowly, were also excluded. To do that, a linear mixed-effects model containing only participants and items as random effects and Trial Number as fixed effect was run. The residuals of this atheoretical model were computed. Based on visual inspection of a residual plot, 19 trials (0.5%) in the test phase (with residuals either below  $-1300$  or above  $3200$  ms) were excluded.

### Statistical testing

Linear mixed-effects models were used to analyze the click responses (accuracy<sup>1</sup> and RT<sup>2</sup>) and the eye movement<sup>3</sup> data on the experimental trials (the /b/-targets and the CVC-targets). To account for the categorical nature of the accuracy data, we used a logistic regression model for these data (cf. Dixon, 2008; Jaeger, 2008). The eye-tracking data were transformed into fixation proportions using the empirical logit function. Participants and Items were entered in the model as random factors including random slopes for Items. Group served as fixed effect. The segmental reduction condition (/b/-words) and the syllabic reduction condition (CVC-words) were analyzed independently. This is because a comparison between these two word sets is difficult: Both had to conform to different phonological constraints and could hence

<sup>1</sup>lmer(Accuracy ~ Group + scale(Trial\_Number, scale = F) + (1 | Participant) + (1 + Group | Item), data = test, subset = Trial\_Type == "test\_b", family = binomial).

<sup>2</sup>lmer(RT ~ Group + scale(Trial\_Number, scale = F) + (1 | Participant) + (1 + Group | Item), data = test, subset = Trial\_Type == "test\_b").

<sup>3</sup>lmer((empLogit(targetProp, 20) - empLogit(compProp, 20)) ~ Group + (1 | Participant) + (1 + Group | Item), data = test, subset = Trial\_Type == "test\_b").

**Table 2 | RTs in ms in the test phases of Experiments 1 and 2 for clicks on targets and competitors.**

	RT in ms	Segmental reduction group	Syllabic reduction group	No reduction group
Experiment 1	Mean	1695	1769	1714
	SD	687	803	751
	Min	455	294	555
	Max	7303	8647	9107
Experiment 2	Mean	1898	1880	1730
	SD	1597	868	660
	Min	346	512	455
	Max	34467	9863	6384

not be balanced on other variables (such as word length, lexical frequency, etc.). We therefore focus on the comparison of how the different groups recognize each word set independently (a one-factorial design with three levels: exposed to /b/-reductions, exposed to vowel-deletions, and not exposed to reductions). Trial Number was entered as another fixed effect with values centered around zero in the models for the accuracy and RT data. This variable was added to account for additional variance, as task performance often improves over the course of an experiment. The results for Trial Number, however, will not be reported below. Thus, we tested whether RTs, accuracy scores and target preference (as determined by the difference between proportion of target and competitor fixations) for the reduced words were influenced by the fixed effect of Group. That is, we examine whether the groups differ in how fast and accurately they recognize the reduced /b/-words and the vowel-deleted words and whether they show different target-competitor preferences when they process reduced words. The control group was always mapped on the intercept, so that the analysis gives two regression weights for the factor Group, one for the difference between the control group and the segmental reduction group and one for the difference between the control group and the syllabic reduction group. For the eye-tracking analyses, we had no a priori expectations about when effects would occur. We therefore analyzed the fixation data at all time points, using sliding 200 ms time windows from 200 to 1500 ms after target onset starting at every 100 ms.

### Test phase

**Reaction time data.** Figure 2A displays the mean RTs of all three groups for the reduced /b/-words (visual /b/-targets) and the vowel-deleted words (visual CVC-targets) in the test phase of Experiment 1. In the segmental reduction condition (/b/-targets), all three groups responded about equally fast and no significant differences between the groups emerged ( $b_{\text{Segmental reduction group}} = -17.9$ ,  $SE = 87.5$ ,  $t = -0.2$ ,  $p = 0.84$ ;  $b_{\text{Syllabic reduction group}} = 117.3$ ,  $SE = 87.4$ ,  $t = 1.3$ ,  $p = 0.21$ ). In the syllabic reduction condition (CVC-targets), there was also no main effect of Group ( $b_{\text{Segmental reduction group}} = -32.7$ ,  $SE = 98.9$ ,  $t = -0.3$ ,  $p = 0.77$ ;  $b_{\text{Syllabic reduction group}} = 1.7$ ,  $SE = 97.4$ ,  $t = 0.02$ ,  $p = 0.98$ ). That is, neither of the experimental groups responded faster than the control group to the reduced words. We thus did not observe any adaptation effects in the RT data.

**Accuracy data.** The accuracy data in the test phase of Experiment 1 are displayed in Figure 3A in terms of percentage of correct click responses and SEs. In the segmental reduction condition (visual /b/-targets), the main effect of Group was significant. Both the segmental reduction group ( $b_{\text{Segmental reduction group}} = 3.4$ ,  $SE = 0.7$ ,  $p < 0.001$ ) and the syllabic reduction group ( $b_{\text{Syllabic reduction group}} = 2.3$ ,  $SE = 0.5$ ,  $p < 0.001$ ) gave more correct responses to /b/-targets than the control group. We thus observed an adaptation effect for both experimental groups in the accuracy data for the segmental reductions.

For the syllabic reductions (visual CVC-targets), the main effect of Group was not significant ( $b_{\text{Segmental reduction group}} = 0.2$ ,  $SE = 0.3$ ,  $p = 0.52$ ;  $b_{\text{Syllabic reduction group}} = 0.3$ ,  $SE = 0.3$ ,  $p = 0.26$ ). That is, neither of the experimental groups differed from the control group. We thus did not observe a significant adaptation effect for either group.

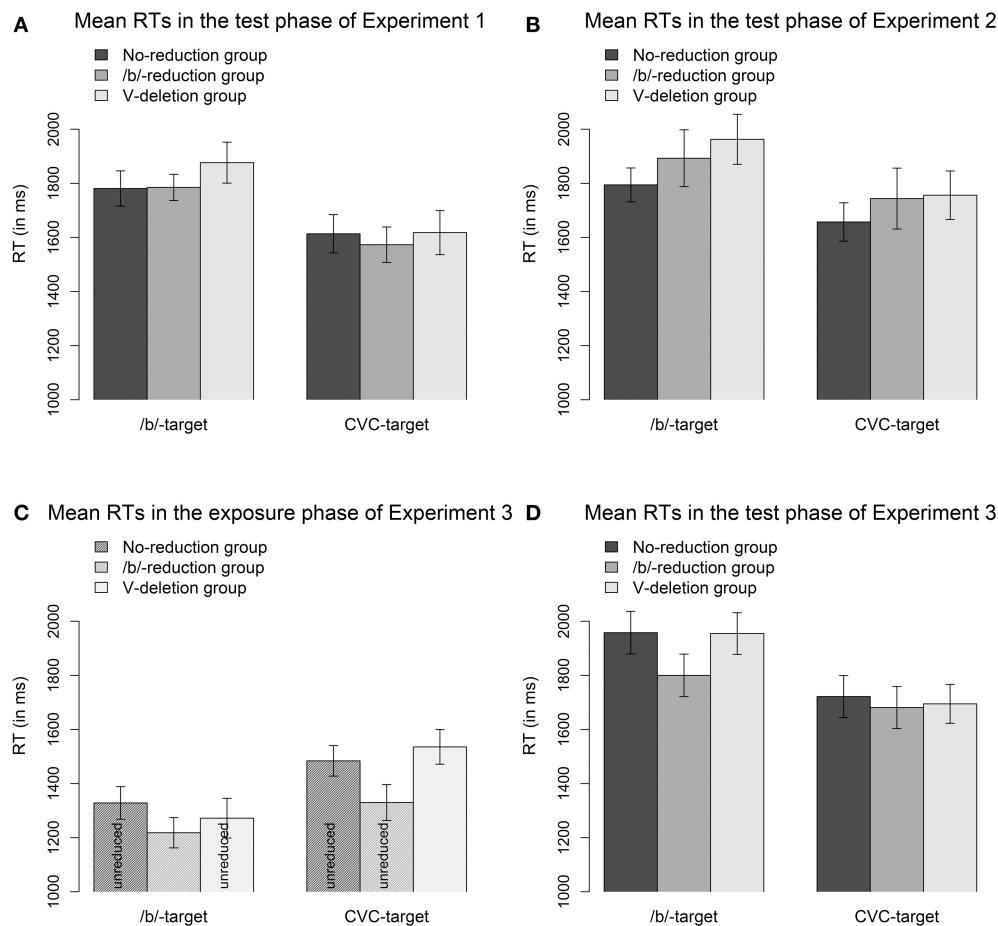
**Eye movement data.** The eye movement patterns for the segmental reduction condition (visual /b/-targets) of the two experimental groups compared to the no-reduction control group are displayed in Figures 4A,B. Early on, in a descriptive time window from 200 to 500 ms after target onset, the control group (represented by black lines) looks more often to the competitors (dashed lines) when hearing a reduced /b/-word than the segmental reduction group (in red, Figure 4A) or the syllabic reduction group (in green, Figure 4B). From around 500 ms onwards, all three groups show a similar preference for the /b/-targets (solid lines).

Statistical analyses considered time windows of 200 ms length which started at 200 ms after target onset and were then shifted by 100 ms (i.e., the following time windows were analyzed: 200–400, 300–500, 400–600, ..., 1300–1500 ms). In the following and both subsequent experiments, only time windows showing significant effects are reported. If several consecutive 200 ms time windows were significant (e.g., the time windows 200–400 and 300–500 ms), the values reported are those for the accumulated time window.

The difference in target-competitor preference between the segmental reduction group and the control group did not reach significance. The main effect of Group, however, was marginally significant for the syllabic reduction group in the time window from 300 to 500 ms after target onset ( $b_{\text{Syllabic reduction group}} = 0.6$ ,  $SE = 0.3$ ,  $t = 1.9$ ,  $p = 0.06$ ). That is, we observed a weak adaptation effect for the syllabic reduction group in the segmental reduction condition, hence a weak generalization of learning across reduction types.

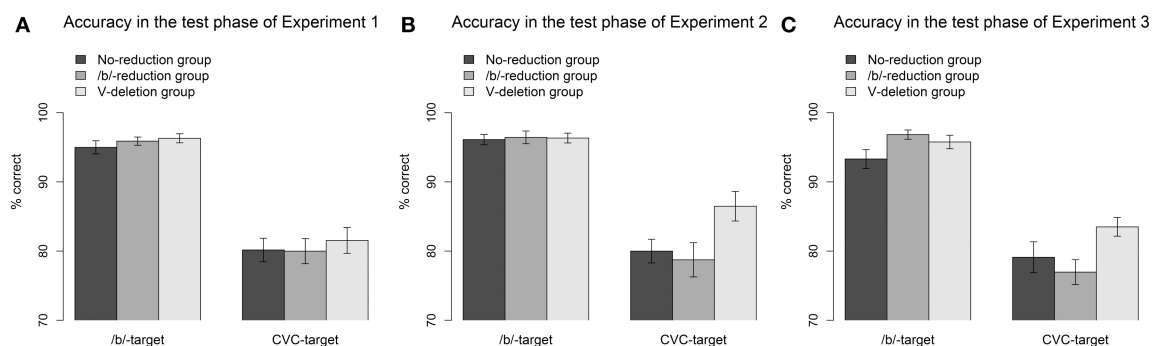
Figures 4C,D display the corresponding eye movement data for the syllabic reduction condition (visual CVC-targets). In the first 900 ms after target onset, all three groups show a very similar pattern for the vowel-deleted words. Only later, the two experimental groups have descriptively a greater target preference for the CVC-targets than the control group.

Statistical analyses did not reveal a significant difference between the control group and the segmental reduction group, but revealed that the main effect of Group was significant in the time window from 1100 to 1400 ms for the syllabic reduction group ( $b_{\text{Syllabic reduction group}} = 0.9$ ,  $SE = 0.4$ ,  $t = 2.2$ ,  $p < 0.05$ ).



**FIGURE 2 | Mean RTs and SEs in the test phases of Experiment 1 (A), Experiment 2 (B), and Experiment 3 (D) and in the exposure phase of Experiment 3 (C).** In the test phases, the /b/-words and CVC-words were

reduced for all groups. In the exposure phases, the /b/-words were reduced only for the /b/-reduction group (segmental reduction group) and the CVC-words were reduced only for the V-deletion group (syllabic reduction group).



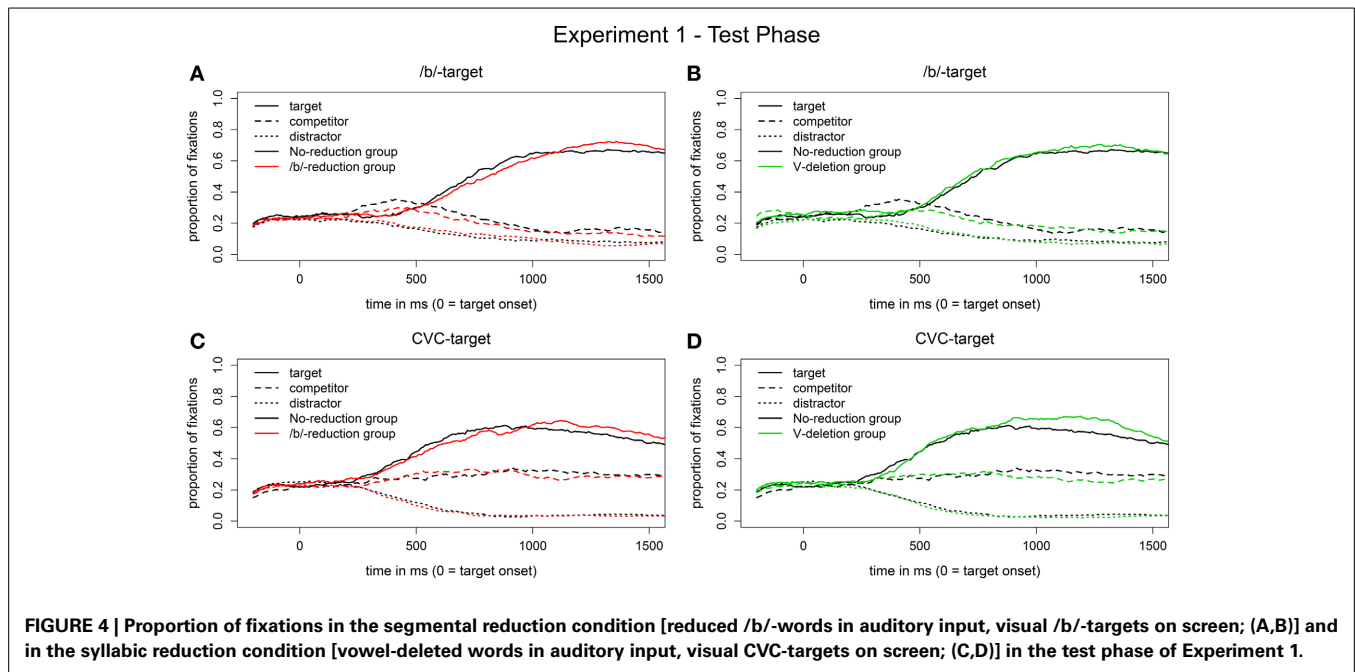
**FIGURE 3 | Accuracy in % correct click responses and SEs for the reduced /b/-words (visual /b/-targets) and the vowel-deleted words (visual CVC-targets) in the test phases of Experiment 1 (A), Experiment 2 (B), and Experiment 3 (C).**

In this time window, the syllabic reduction group had a greater target-competitor preference for the CVC-words than the control group. For the syllabic reduction group, we thus found an adaptation effect.

## DISCUSSION

In Experiment 1, we found adaptation effects for both the segmental and the syllabic reductions. Learning about segmental reductions was evident in the accuracy data but not in the





eye-tracking data. For the syllabic reductions, this pattern was reversed: A learning effect was found in the eye-tracking data but not in the accuracy data. Moreover, there was also evidence of generalization of learning across reduction types. Generalization across reduction types, however, was only found in one direction: learning about vowel deletions generalized to /b/-reductions, as shown by the accuracy data and the eye movement data for the segmental reductions. In contrast, learning about /b/-reductions did not generalize. That is, the segmental reduction group could not apply their experience with reductions to the vowel-deleted words.

The learning effects found in Experiment 1 seem somewhat weak. An explanation for this may be that the potentially reduced words in the exposure phase were not highly predictable. Participants did not see the potentially reduced words on the computer screen during the exposure phase and these words appeared early in the sentences, which were in fact designed to predict the targets (e.g., in *Pas in een [b]/[m]inderij wordt een boek of tijdschrift afgemaakt*, the target *tijdschrift* is predictable and the potentially reduced word *[b]/[m]inderij* is not). Participants may therefore not have been able to predict potentially reduced words. Having information about the upcoming reduced words in advance could however facilitate learning. Jesse and McQueen (2011) found that adaptation to ambiguous fricatives did not take place if those fricatives occurred at the onset of a word presented in isolation. They concluded that lexical information likely has to be available when the ambiguous sound is initially being processed. The present study investigates adaptation to another form of deviation, which also occurs at the beginning of the words. Predictable sentence contexts may provide sufficient cues about the upcoming words so that adaptation may be possible. Experiment 2 was run to test this hypothesis.

## EXPERIMENT 2

Experiment 2 tested whether providing additional information about the reduced words in the exposure phase might strengthen the learning effects found in Experiment 1. Therefore, we changed the exposure sentences for the experimental words, leaving the filler sentences for the /m/-words and the consonant-cluster words intact. The sentence contexts now predicted the potentially reduced words. To avoid the orthographic versions of the reduced words appearing twice on the screen, the clicking task was not used in the exposure phase. Instead, participants simply listened to the exposure sentences and were asked to answer questions about the content of some of the filler sentences (those containing /m/- or CC-words).

The test phase was kept the same as in Experiment 1, apart from minor changes in three sentences (see Methods section). Further purposes of Experiment 2 were to replicate the generalization effect from vowel-deleted words to reduced /b/-words found in Experiment 1 and to test whether, with predictable sentences, a generalization effect in the other direction (from reduced /b/-words to vowel deletions) might occur.

## METHODS

### Participants

Sixty Dutch participants of the Max Planck Institute's subject pool, none of whom had participated in Experiment 1, were paid for their participation. All had normal hearing and normal or corrected-to-normal vision.

### Design

The design was similar to that in Experiment 1. The main difference was a change in task during the exposure phase, where participants had to answer questions regarding the content of some

of the reduction-free sentences without their eye movements being tracked.

### Materials

As in Experiment 1, the exposure and the test phases consisted each of 96 trials (48 experimental trials containing either /b/-words or CVC-words and 48 filler trials containing either /m/-words or CC-words). While for the fillers the same exposure sentences as in Experiment 1 were used, new exposure sentences were generated for the experimental conditions (the potentially reduced /b/-words and the vowel-deleted words). The critical words now appeared toward the end of the sentences (e.g., *Als een manuscript gedrukt is, moet het naar de [b]/[m]inderij*. “When a manuscript is printed, it has to go to a book binder”) and were predicted by the semantic context (see cloze test below). The materials for the test phase were taken from Experiment 1. Only three target words were changed slightly (*bankier* “banker” → *bankiers* “bankers,” *benauwen* “to oppress” → *benauwd* “sultry,” *coulisse* “wing [of theater stage]” → *coulissen* “wings, pl.”) so that it was possible to create more natural sentences for the exposure phase.

### Cloze tests

Cloze tests were run to check the degree of predictability of the potentially reduced words in the exposure sentences. The 48 sentences were presented in a randomized order with the critical word replaced by a gap. Participants were instructed to complete these sentences with one word. They were asked to type in at least one answer but had the possibility to give up to seven. After typing in their answer(s), participants saw the same sentence again completed with the corresponding /b/- or CVC-target. They were asked to rate how well the proposed solution completed the sentence context on a scale from 1 (“Word does not fit at all”) to 7 (“Word fits perfectly”). The cloze tests were self-paced; it took participants 15 to 30 min.

An initial test with eighteen Dutch native speakers of the Max Planck Institute’s subject pool, who had not participated in Experiment 1, showed that for some sentences the target word was mentioned in less than 25% of cases. These were improved if possible. A second version of the cloze test was run with 19 new Dutch participants. We analyzed the percentages of mentioned target words in the sentence completion task and the mean ratings for the targets in the rating task. The critical /b/-words were mentioned in 36% of the cases, while the critical CVC-words were mentioned in 51% of the cases. This difference does not reflect a frequency effect, as the /b/-targets are more frequent than the CVC-targets (see Tables S1, S2). But it can possibly be explained by the higher constraints on the initial selection of the /b/-words. Only /b/-words were chosen which had a nasal in third position and for which a /m/-initial competitor with as much onset overlap as possible existed. Similar constraints on the CVC-words were less strong, as the consonants in first and third position could vary. Although participants did not come up with our solutions in many cases, they rated those solutions very highly on average: On a scale from 1 to 7, with higher ratings meaning better fits, participants rated the /b/-targets 6.1 and the CVC-targets 6.3 on average.

### Stimulus construction

The new exposure sentences were recorded by the same female Dutch speaker who provided the stimuli for Experiment 1. The reduced stimuli were created in the same way as described in Experiment 1.

### Procedure

Participants were tested in a sound-proof booth. They were told that the experiment consisted of two parts. For the first part, they were asked to listen to sentences that were presented over headphones and to answer questions regarding the content of these sentences (by clicking on one out of two suggested solutions) that might appear at random points in time on the screen.

Each exposure sentence was preceded by 500 ms of silence and followed by 2000 ms of silence. If a question and two possible solutions were to appear on the screen (after six /m/-word sentences and after six CC-word sentences, i.e., in  $1/8$ th of the exposure trials), they followed the auditory stimulus immediately. After participants had clicked on the screen, it took 1000 ms before the next exposure trial started. The order in which the exposure sentences were played was randomized for each participant individually. Participants had the opportunity to take a break approximately halfway through the experiment, after the 50th stimulus (out of 96).

The procedure of the test phase was identical to the one in Experiment 1, except that eye movements were monitored using an SR Research EyeLink II, sampling at 500 Hz. An experimental session took approximately 30 min.

## RESULTS

### Exclusion criteria

The same criteria as in Experiment 1 were applied for trial exclusion. This led to the exclusion of 2.2% of the data due to fixations outside of the screen area and of another 1.3% due to failure to click on the target or the potentially confusable competitor. An additional 0.5% of trials were discarded because they were considered to be RT outliers (with residual values either below −2300 or above 3100 ms). For the eye-tracking data, we analyzed the data from the better eye of the participants (i.e., the eye that showed less error in the validation of the calibration of the eye-tracker).

### Exposure phase

Participants of all groups hardly made errors in the comprehension questions of the exposure phase. Each group obtained a score of 99% correct responses.

### Test phase

**Reaction time data.** The lower part of Table 2 shows the descriptive statistics for the RT data in the test phase of Experiment 2. The mean RTs and their SEs of all three groups for the reduced /b/-words (visual /b/-targets) and vowel-deleted words (visual CVC-targets) are displayed in Figure 2B. The no-reduction control group seems to respond slightly faster than the two experimental groups in both the segmental reduction condition (/b/-targets) and the syllabic reduction condition (CVC-targets). However, the main effect of Group was not significant in either condition (/b/-targets:  $b_{\text{Segmental reduction group}} = 118.2$ ,  $SE = 127.0$ ,  $t = 0.9$ ,  $p = 0.38$ ;  $b_{\text{Syllabic reduction group}} = 191.2$ ,  $SE = 130.2$ ,

$t = 1.5$ ,  $p = 0.15$ ; CVC-targets:  $b_{\text{Segmental reduction group}} = 93.0$ ,  $SE = 121.5$ ,  $t = 0.8$ ,  $p = 0.43$ ;  $b_{\text{Syllabic reduction group}} = 123.2$ ,  $SE = 121.5$ ,  $t = 1.0$ ,  $p = 0.33$ ). As there was no main effect of Group in the RT data indicating that one or both of the experimental groups responded faster to the reduced targets than the control group, we did not observe any adaptation effect.

**Accuracy data.** Figure 3B shows the accuracy data in percentages correct responses and SEs of all three groups for the reduced /b/-words (visual /b/-targets) and vowel-deleted words (visual CVC-targets). All three groups performed near ceiling in the segmental reduction condition (/b/-targets). There was no difference between the groups ( $b_{\text{Segmental reduction group}} = -0.4$ ,  $SE = 0.5$ ,  $p = 0.44$ ;  $b_{\text{Syllabic reduction group}} = -0.6$ ,  $SE = 0.5$ ,  $p = 0.22$ ) indicating that the experimental groups did not respond more accurately than the control group. We thus did not observe an adaptation effect in the accuracy data for the segmental reduction condition.

In the syllabic reduction condition (CVC-targets), the main effect of Group was significant for the syllabic reduction group ( $b_{\text{Syllabic reduction group}} = 0.9$ ,  $SE = 0.3$ ,  $p < 0.01$ ) but not for the segmental reduction group ( $b_{\text{Segmental reduction group}} = 0.2$ ,  $SE = 0.3$ ,  $p = 0.54$ ). That is, only the syllabic reduction group gave more correct answers when hearing a vowel-deleted word than the no-reduction control group. We thus observed a learning effect for the syllabic reduction group, but no generalized learning effect for the segmental reduction group.

**Eye movement data.** Figures 5A,B shows the eye-movement patterns in the segmental reduction condition (visual /b/-targets) for the segmental reduction group (in red) and the syllabic reduction group (in green) compared to the no-reduction control group (in black). All three groups behave very similarly when hearing reduced /b/-words. There was indeed no main effect of Group. That is, we did not observe a learning effect for the segmental reduction condition in the eye-tracking data.

The corresponding eye movement data for the syllabic reduction condition (visual CVC-targets) are displayed in Figures 5C,D. Statistical analysis revealed a marginal main effect of Group ( $b_{\text{Segmental reduction group}} = -0.6$ ,  $SE = 0.3$ ,  $t = -2.0$ ,  $p = 0.06$ ) in the time window from 200 to 500 ms after target onset. The segmental reduction group had a smaller preference for the CVC-targets over the CC-competitors than the control group in this time window. We thus observed a marginal inhibitory effect for the segmental reduction group, given that participants in this group, who had experience with another type of reduction, showed a smaller target preference than participants in the control group, who had not been exposed to any reductions. Furthermore, no learning effect was found for the syllabic reduction group.

## DISCUSSION

Experiment 2 was conducted to replicate the findings of Experiment 1 and to test whether predictability of the reduced words during exposure enhances the learning effects. As in Experiment 1, adaptation was observed in the syllabic reduction condition. Contrary to the previous experiment, it was found

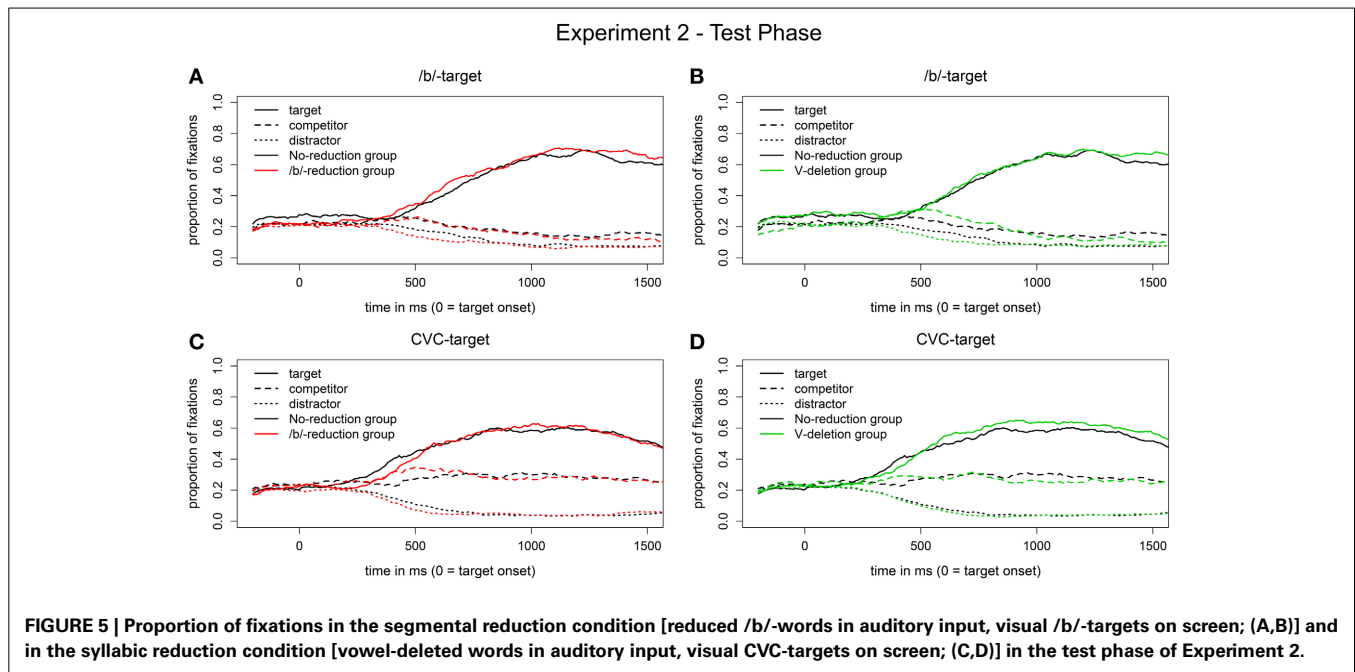
in the accuracy data, not in the eye-tracking data. The pattern of target and competitor fixations for the syllabic reduction group, however, was in the expected direction (see Figure 5D). We did not replicate the learning effect for segmental reductions found in the accuracy data in Experiment 1. Neither could we replicate the generalized learning effect for the syllabic reduction group for vowel-deletions to /b/-reductions (that was also evident in the accuracy data of Experiment 1). Another generalization effect emerged, however. In contrast to Experiment 1, the segmental reduction group differed from the control group when dealing with vowel deletions. In the eye-tracking data, they showed a smaller target-competitor preference for CVC-targets. That is, even though they did not show a learning effect for /b/-reductions, participants in the segmental reduction group seemed to be hindered by their exposure to /b/-reductions and struggled more with recognizing the vowel-deleted words than the control group.

In Experiments 1 and 2, we found learning effects for repeated /b/-reductions and vowel-deletions. At this point, we cannot say whether these effects are truly word-specific, meaning that they arose because the reduced forms were stored after their first encounter in the mental lexicon and then accessed again as they were encountered the second time in the test phase. The observed effects could also have arisen because of rule abstraction. To determine which mechanism is responsible for the learning effects found for repeated reduced words in Experiments 1 and 2, we tested whether learning can generalize to other words of the same reduction type in Experiment 3. If there is no or only weak evidence for generalized learning, then the effects found for repeated words are very likely to be word-specific. In contrast, if there is strong evidence for generalized learning, then the effects found for repeated words are likely due to abstraction processes.

The null result for the segmental reductions in Experiment 2 suggests that predictable sentences alone might not be enough to induce a stable adaptation effect. In Experiment 3, we therefore combined aspects of the exposure phase of Experiment 1 (eye-tracking with printed words on the screen) with aspects from Experiment 2 (predictable sentence context). This procedure should render the reduced target words highly predictable, which in turn could lead to a strong learning effect. Using eye-tracking in the exposure phase can tell us whether participants actually make use of the sentence context (i.e., they might already look at the target word before it is mentioned).

## EXPERIMENT 3

In Experiment 3, we tested whether learning about reductions can generalize across words (within a reduction type). To that end, new /b/-words and new CVC-words were selected for the exposure phase and new exposure sentences were created in which those words were predictable. In the exposure phase, participants had to click on the potentially reduced /b/-target and CVC-target words, while their eye-movements were recorded. The test phase was the same as in Experiment 1. Importantly, the target words used in the test phase did not occur in the exposure phase. Apart from the generalization of learning within a reduction type, Experiment 3 again tests generalization of learning across



reduction types and aims to replicate and extend the results from Experiment 1 on this issue.

## METHODS

### Participants

Sixty Dutch participants of the Max Planck Institute's subject pool, none of whom had participated in the previous experiments, took part for a small remuneration. All reported normal hearing and normal or corrected-to-normal vision.

### Design

The design was very similar to the one of Experiment 1, except for changes in the exposure phase. Predictable exposure sentences were created for new potentially reduced /b/-words and vowel-deleted words which served as target words in an eye-tracking paradigm. That is, participants had to click on the orthographic form of these words while their eye movements were recorded. The test phase was the same as in Experiment 1. Due to the changes in the exposure phase, the targets in the test phase were new to participants and not repeated as in Experiments 1 and 2 (see Table 3).

### Materials

As in Experiments 1 and 2, the exposure and the test phases consisted each of 96 trials (48 experimental trials containing either /b/-words or CVC-words and 48 filler trials containing either /m/-words or CC-words). The exposure sentences for the /m/-words and the CC-words were the same as in Experiment 1. The exposure sentences for the potentially reduced /b/-words and vowel-deleted words were constructed anew. These critical words appeared again toward the end of the sentences and were predicted by the semantic context.

For the selection of the 24 exposure /b/-targets and the 24 exposure CVC-targets, the same constraints applied as for the

respective targets of the test phase. The criteria for the selection of their "competitors" were less strict. These only overlapped in the initial consonantal part for reduced forms, but were additionally matched on word class (e.g., *bandiet* "bandit" would be reduced to [mandit] and would compete for recognition with [mirakəl] *mirakel* "miracle"; *kanaal* "canal" would be reduced to [kna] and would compete with [knɛxt] *knecht* "servant"). The materials for the test phase were taken from Experiment 1.

### Stimulus construction

The new exposure sentences were recorded by the same female Dutch speaker as in Experiments 1 and 2. The reduced stimuli were created as described in Experiment 1.

### Procedure

The procedure was similar to Experiment 1 except for changes in the exposure phase, in which participants had to click on the potentially reduced word that was predictable from the sentence context. The computer display always showed a /b/-word, a /m/-word, a CVC-word and a consonant-cluster word on the screen. Exposure and test displays differed only in the phonological similarity of target and competitor words which were more similar in the test phase (e.g., exposure trial: *bandiet* vs. *mirakel*; test trial: *binderij* vs. *minderjarig*). An experimental session took approximately 25 min.

## RESULTS

### Exclusion criteria

Trials were excluded based on the same criteria as used in Experiments 1 and 2. Due to fixations outside of the screen, 2.6% of the trials were removed. Another 0.6% were discarded due to failure to click on the target or the potentially confusable competitor. Fifteen trials (0.3%) in the exposure phase (with residuals either below -1100 or above 2500 ms) and 29 trials (0.5%) in the

**Table 3 | Experimental design and types of stimuli in Experiment 3.**

	Trial type	Canonical word-form	Segmental reduction group	Syllabic reduction group	Control group
			/b/ → [m]	Full vowel deletion	No reduction
Exposure phase	Experimental	/bandit/	<b>[m</b> andit]	[bandit]	[bandit]
	Filler	/matros/	[matros]	[matros]	[matros]
	Experimental	/kanal/	[kanal]	<b>[kna]</b> l]	[kanal]
	Filler	/knɔflok/	[knɔflok]	[knɔflok]	[knɔflok]
Test phase	Experimental	/bɪndərəɪ/		<b>[m</b> ɪndərəɪ]	
	Filler	/muras/		[muras]	
	Experimental	/parat/		<b>[pra]</b> t]	
	Filler	/xlans/		[xlans]	

Reduced segments are marked bold. The potentially reduced /b/-initial words and the vowel-deleted words of the exposure phase were not repeated in the test phase.

test phase (with residuals either below −1700 or above 2800 ms) were considered to be RT outliers and hence excluded. For the eye-tracking results, the data of the participants' right eye were analyzed.

An overview of the accuracy data in the exposure and test phases can be found in **Table 4**. In the exposure phase, practically no errors were made. In the test phase, we again observe a high percentage of errors for the vowel-deleted words in all three groups. **Table 5** displays the descriptive statistics for the RT data in the exposure and test phases. All three groups took longer to give a click response in the test phase (where the sentence context was neutral and the words on the screen were quite similar to each other) than in the exposure phase (where they could use the sentence context to predict the target word). The negative minima for RTs in the exposure phase confirm that the target words in this phase were indeed predictable, as some participants responded even before target onset.

### Exposure phase

The /b/-words were reduced only for the segmental reduction group and the CVC-words were reduced only for the syllabic reduction group. There were virtually no errors (see **Table 4**). Moreover, in contrast to previous results (Poellmann et al., under revision), there was no consistent effect of reduction, neither in RTs nor in the eye-tracking data (see **Figures 2C, 6**; the main effect of Group was not significant for the groups who heard reduced forms indicating that they did not have more difficulties in recognizing the targets than the control group). The data from the exposure phase reflect that the target words were predictable, as participants in all groups already showed a preference for the target before it was mentioned (see the time windows from −200 to 0 ms in **Figure 6**). Apparently, the words in the exposure phase were recognized efficiently whether they were reduced or not.

### Test phase

**Reaction time data.** **Figure 2D** displays the mean RTs and SEs of all three groups for the reduced /b/-words (visual /b/-words) and the vowel-deleted words (visual CVC-words) in the test phase of Experiment 3. The segmental reduction group seems to respond somewhat faster to the reduced /b/-words, while all three

groups seem to respond about equally fast to the vowel-deleted words.

Statistical analyses did not show a main effect of Group—neither in the segmental reduction condition (/b/-targets:  $b_{\text{Segmental reduction group}} = -177.1$ ,  $SE = 108.5$ ,  $t = -1.6$ ,  $p = 0.12$ ;  $b_{\text{Syllabic reduction group}} = -15.2$ ,  $SE = 106.4$ ,  $t = -0.1$ ,  $p = 0.92$ ) nor in the syllabic reduction condition (CVC-targets:  $b_{\text{Segmental reduction group}} = -7.8$ ,  $SE = 102.8$ ,  $t = -0.1$ ,  $p = 0.92$ ;  $b_{\text{Syllabic reduction group}} = 11.7$ ,  $SE = 103.0$ ,  $t = 0.1$ ,  $p = 0.92$ )—indicating that all groups responded equally fast to both types of target words. That is, the groups experienced with reduced forms did not respond faster than the less experienced control group. We thus did not observe any adaptation in the RT data.

**Accuracy data.** The accuracy data in terms of percentage correct responses and their SEs of all groups can be found in **Figure 3C**. Both experimental groups seem to give more accurate responses to reduced /b/-words (visual /b/-words) than the control group. For the vowel-deleted words (visual CVC-words), only the syllabic reduction group seems to respond more accurately than the control group. This, however, was not confirmed by statistical analyses. The main effect of Group was not significant either in the segmental reduction condition (/b/-targets:  $b_{\text{Segmental reduction group}} = 0.8$ ,  $SE = 0.5$ ,  $p = 0.16$ ;  $b_{\text{Syllabic reduction group}} = 0.1$ ,  $SE = 0.5$ ,  $p = 0.90$ ) or in the syllabic reduction condition (CVC-targets:  $b_{\text{Segmental reduction group}} = -0.4$ ,  $SE = 0.3$ ,  $p = 0.16$ ;  $b_{\text{Syllabic reduction group}} = 0.4$ ,  $SE = 0.3$ ,  $p = 0.21$ ). That is, neither of the experimental groups gave more correct answers to the reduced targets than the control group. We thus did not observe any adaptation effects in the accuracy data.

**Eye movement data.** **Figures 7A,B** display the eye movement pattern of the two experimental groups plotted against the patterns of the control group (in black) for the segmental reduction condition. Both experimental groups show a greater preference for the target over the competitor for the reduced /b/-words than the control group, descriptively from around 700 ms onwards (when the colored lines diverge from the black lines). This difference is bigger for the segmental reduction group (in red).

The main effect of Group reached significance only for the segmental reduction group ( $b_{\text{Segmental reduction Group}} = 1.0$ ,

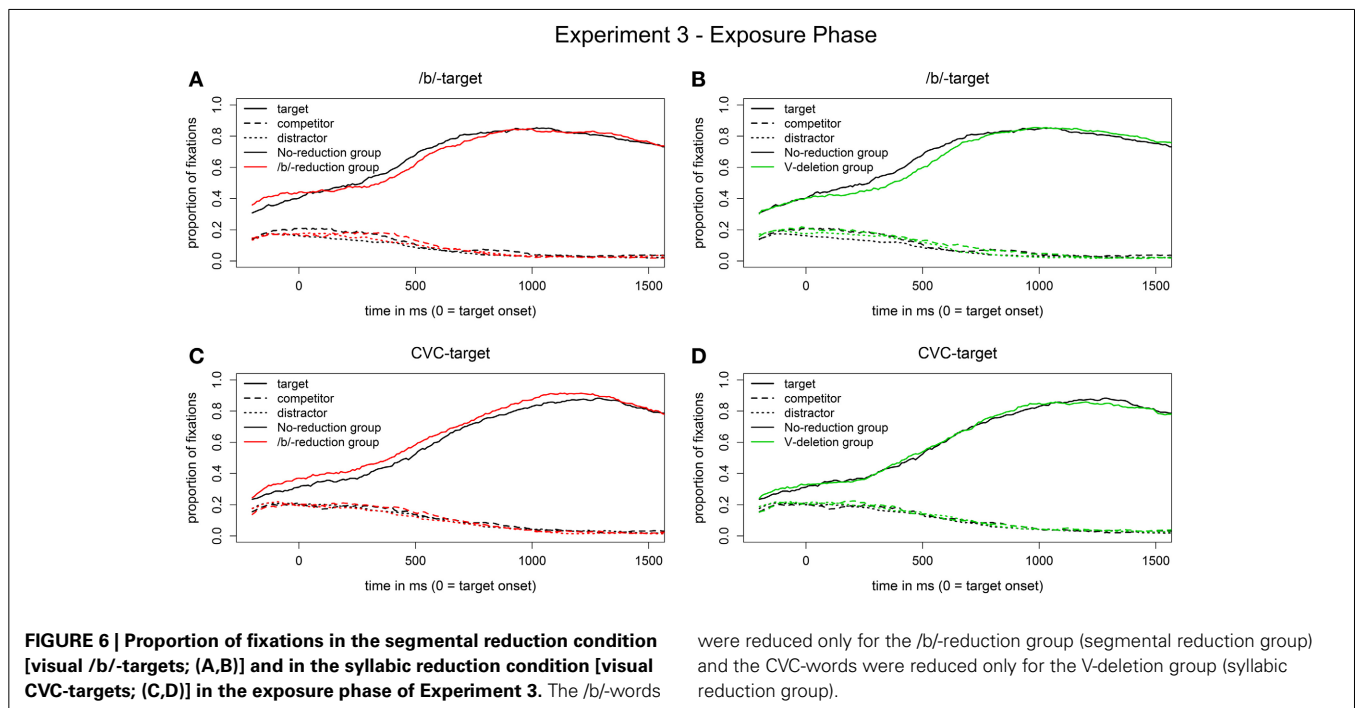


**Table 4 | Accuracy data of the exposure and test phases of Experiment 3.**

% Click responses		Segmental reduction group			Syllabic reduction group			No reduction group		
		Target	Comp.	Distr.	Target	Comp.	Distr.	Target	Comp.	Distr.
Exposure	/b/-word	99.4	0.0	0.0	99.4	0.0	0.0	99.6	0.0	0.0
	CVC-word	99.6	0.0	0.0	99.4	0.0	0.4	99.8	0.0	0.0
Test	/b/-word	95.8	3.1	0.0	95.2	4.2	0.0	92.3	6.7	0.2
	CVC-word	75.8	22.7	0.0	83.3	16.5	0.0	79.0	20.8	0.0

**Table 5 | RT in ms in the exposure and test phases of Experiment 3 for clicks on targets and competitors.**

RT in ms	Segmental reduction group		Syllabic reduction group		No reduction group	
	Exposure	Test	Exposure	Test	Exposure	Test
Mean	1280	1754	1429	1876	1418	1889
SD	610	736	761	947	644	955
Min	−40	554	−40	430	−39	677
Max	7374	8015	7060	10854	6419	9599

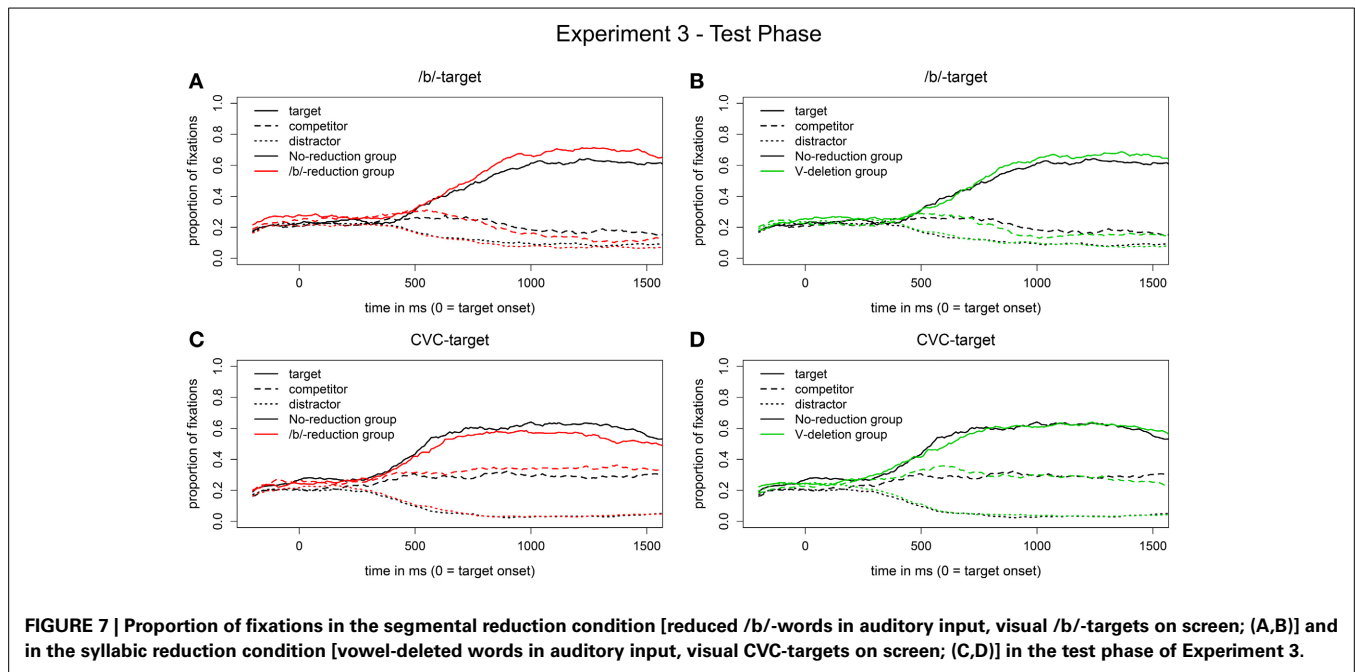


$SE = 0.4$ ,  $t = 2.4$ ,  $p < 0.05$ ) from 1100 ms onwards. That is, the segmental reduction group, but not the syllabic reduction group, outperformed the control group on the reduced /b/-words. We thus observed a within-reduction-type generalization effect in the segmental reduction condition.

The corresponding eye-movement data for the syllabic reduction condition are displayed in **Figures 7C,D**. The segmental reduction group (in red) shows a smaller target-competitor preference for the CVC-targets than the control group, descriptively from 500 ms onwards. The syllabic reduction group

shows a similar pattern from 500 to 700 ms after target onset.

Statistical analyses showed a marginally significant main effect of Group only for the segmental reduction group ( $b_{\text{segmental reduction group}} = -0.9$ ,  $SE = 0.5$ ,  $t = -2.0$ ,  $p = 0.06$ ) in the time window from 1200 to 1500 ms. That is, the segmental reduction group but not the syllabic reduction group had a significantly smaller target preference for the vowel-deleted words than the control group. We thus did not observe a learning effect for the syllabic reduction group and found a marginal inhibitory effect



for the segmental reduction group. Participants in the latter group seem to be hindered by their prior exposure to another reduction type.

## DISCUSSION

The aim of Experiment 3 was to test whether learning about reductions can generalize within and across reduction types. In the exposure phase, listeners were provided with predictive sentence contexts and with orthographic information about the critical words, as they saw the orthographic forms of the potentially reduced words on the computer screen. The results from the exposure phase did not show any effects of reduction. That is, neither the segmental reduction group nor the syllabic reduction group were slowed down or had a smaller target preference when hearing reduced forms. This is very likely due to the predictive sentence context. Participants were already expecting the target and looking at it before it was actually mentioned. Hearing it then in reduced form did not disturb the recognition process any more. Note that these data apparently are in contrast with the data of Brouwer et al. (2013), who found that even predictable words suffer from reduction costs. The difference, however, might be due to the stimulus material, with our material being constructed to allow prediction of the target word, while Brouwer et al. used materials from a speech corpus. For reduced words which were particularly predictable, they also observed less reduction costs.

In the test phase, we found clear evidence for generalization of learning within reduction type for the segmental reduction group in the eye-tracking data. No such generalization effect was found for the syllabic reduction group. Contrary to the word-specific learning effects found in Experiments 1 and 2, the within-reduction-type generalizations were stronger for /b/-reductions than for vowel-deletions.

As for generalization of learning across reduction types, we did not replicate the transfer of learning from vowel-deletions

to /b/-reductions for the syllabic reduction group found in Experiment 1. There was a trend going in this direction though (see **Figure 7B**). However, we replicated the marginal inhibitory effect of the segmental reduction group found in Experiment 2. That is, the segmental reduction group did not benefit from its exposure to /b/-reductions and had instead slightly greater problems in recognizing vowel-deletions than the no-reduction control group.

## GENERAL DISCUSSION

The present study investigated whether and how listeners can adapt when they encounter reduced word forms. In the introduction, we argued for a continuum of possible adaptation mechanisms that are more or less general. At the specific end, listeners may only adapt to exactly the same words. A more general adaptation would allow generalization to other words of the same or a similar reduction type. Experiments 1 and 2 tested learning effects for repeated segmental and non-morphemic syllabic reductions. Experiment 3 examined whether these learning effects were word-specific by testing whether learning about these reductions generalizes to new words of the same reduction type (within-reduction-type generalization). All three experiments investigated whether experience with one reduction type helps the listener in dealing with another reduction type (across-reduction-type generalization).

Experiments 1 and 2 showed evidence of learning for repeated vowel-deletions but, surprisingly, far less so for repeated /b/-reductions. In contrast, Experiment 3 revealed a strong within-reduction-type generalization effect in the eye-tracking data for the /b/-reductions that was not found for the vowel-deletions. In Experiments 2 and 3, the segmental reduction group further showed a marginal inhibitory effect; they had greater difficulties than the control group dealing with unfamiliar vowel-deletions. Another pattern that was consistently observed (even though not

always statistically significant) was that the syllabic reduction group made fewer errors for both the same and other vowel-deleted words (see **Figures 3A,B,C**, focusing on the CVC-targets, e.g., *paraat* produced as *p'raat*). Next to this reduction-specific adaptation, this group also showed generalization of learning across reduction types (from vowel-deletions to /b/-reductions). This generalization effect, however, could not always be found: It was absent in Experiment 2 where task demands in the exposure phase were low and the predictability of the reduced word was high. It was present in Experiment 1, where task demands in the exposure phase were high, but the predictability of the reduced word was low. Finally, a trend was observed again in Experiment 3, where both task demands and the predictability of the reduced word in the exposure phase were high.

The results of Experiment 3 shed further light on the learning effects found in Experiments 1 and 2. For the segmental reductions, strong generalization of learning to new reduced /b/-words was observed. This suggests that, for the /b/-reductions investigated here, recognition predominantly occurs via abstraction rules. It is therefore likely that abstraction processes also play a role in the recognition of repeated reduced /b/-words. The learning effect found for repeated reduced /b/-words in Experiment 1 thus is very likely not a word-specific adaptation. For the vowel-deletions, no generalization of learning to other vowel-deleted words was observed in Experiment 3. The adaptation effects for repeated vowel-deleted words found in Experiments 1 and 2 are therefore very likely due to storage of these reduced forms and hence are word-specific. Similarly, Hanique et al. (2013) claim that, if the absence of schwa in the prefix of Dutch past participles is due to categorical processes, these schwa-deleted forms are stored in the mental lexicon.

Lexical storage is not only useful if a listener encounters a reduced word for the first time, but may also help to build up abstraction rules for later generalization of learning to other words that show the same reduction pattern. It is therefore surprising that we did not find any benefit for repeated reduced /b/-words in Experiment 2, while we did find a benefit for repeated vowel-deleted words under the same circumstances. Furthermore, although small, such a benefit was found for repeated reduced /b/-words in Experiment 1, where participants were involved in a more active task, but where the reduced /b/-words were hardly predictable. One possible explanation for these findings is based on the difference in saliency between the two reduction types. In the vowel-deletions, an entire segment is completely deleted, whereas in the /b/-reductions the segment is only weakened. The vowel-deletions are thus more striking than the /b/-reductions and potentially are therefore less susceptible to experimental manipulations. Apparently, manipulating the preceding context to make the reduced /b/-words more predictable was not enough to draw participants' attention to that reduction type, while giving listeners a more active task might have achieved this. Learning about reductions might thus only occur if the reduction type is (made) salient enough. Note that in Experiment 3, where learning for /b/-reductions was found, listeners saw the orthographic form of the reduced /b/-words on the screen already in the exposure phase. This may have boosted the learning effect.

The within-reduction-type generalization effect found for new reduced /b/-words in Experiment 3 supports the assumption of an abstractionist mode of lexical access. For the vowel-deletions, only a hint of this generalization effect was observed (in the accuracy data). An important difference between /b/-reductions and vowel-deletions that could explain this discrepancy is input consistency. In the /b/-initial words that were to be reduced, the /b/ was always followed by a vowel and a nasal. The structure of the CVC-words was less consistent: The first consonant could be /k, x, p/, the vowel to be deleted was variable and the second consonant was either a liquid or /n/. The phonological context surrounding the reduced segment and the reduced segment itself varied thus more in the vowel-deletions than in the /b/-reductions. This input variability for vowel-deletions may have been too high for the successful generation of an abstract mapping rule. This very likely restricts generalized learning about syllabic reductions to morphemes that show a high frequency of occurrence across words.

There hence seems to be evidence for two types of adaptation: word-specific adaptation to inconsistent phonological patterns and word non-specific adaptation to consistent patterns. More general learning effects, if observed at all, were marginal. This already suggests that it is hard to apply the knowledge of one reduction type to another in case the two reduction types differ substantially. Nevertheless, we observed such a non-specific adjustment to reductions for the syllabic reduction group. Listeners in this group showed a greater tolerance to /b/-reductions than the control group. Possible factors that likely play a role in this uni-directional facilitative effect are input variability and degree of reduction. These two factors, however, are (necessarily) confounded in the present study. The vowel-deletions are both more variable in their segmental structure and more severely reduced than the /b/-reductions. Similar conditions were present in the study by Brouwer et al. (2012). Brouwer et al. focused on processing at the lexical level and selected reductions which had more onset overlap with another existing word than with their respective canonical form (e.g., the reduced form [pjutər] from *computer* is at the onset more similar to the word *pupil* than to *computer*). As a consequence, their set of reductions contained a large variety of reductions, making it unlikely that listeners could adapt to a specific form of reduction. With this set of varying reductions, they found similar facilitative effects as observed here for the group exposed to variable vowel deletions. They reported that listeners penalized acoustic mismatches between input and canonical form less strongly when listening to (strongly and therefore not regularly) reduced speech.

Instead of also observing facilitation for the segmental reduction group in dealing with vowel-deletions, we found marginal inhibitory effects. After having been exposed to consistently reduced /b/-words, the segmental reduction group did worse on the more strongly reduced vowel-deleted words than the control group. It might thus be that learning about reduction can only generalize to other reduction types that are of the same or a lesser degree of reduction but not to reduction types that show a higher degree of reduction. Another possibility is that

the vowel-deletions differed in too many ways from the /b/-reductions so that it was not possible to adjust the abstract mapping rule for /b/-reductions to accommodate the variable vowel-deletions.

But why did the segmental reduction group actually differ from the control group in dealing with vowel-deletions? It might be the case that participants in the segmental reduction group expected the speaker to produce reductions only in a consistent way and to a specific degree (e.g., weakening of a segment). This might have biased them against other types of variability and the greater deviation from the canonical form that they encountered in the test phase. The control group, in contrast, had not heard any reductions in the exposure phase. In the subsequent test phase, participants in that group suddenly had to deal with many and various reduced forms. As they could not have built up abstract mapping rules, they probably resorted to flexible, non-specific adjustments, like those observed by Brouwer et al. (2012). Finally, the syllabic reduction group was already used to dealing with variable reduced forms. Participants in this group could therefore handle a consistent and less severe reduction type. How well listeners can handle new reduced forms of a different reduction type might thus also depend on listeners' expectations about a speaker's reduction style and, based on that, on the adaptation mechanisms already in use (specific abstraction rules vs. fast perceptual but non-specific adjustments).

What does this series of eye-tracking experiments tell us about possible constraints and the time-course of learning about reductions? Apparently, the reduced forms have to be noticeable, as learning effects were found for less salient reduction types only if the reduced words appeared in orthographic form on the screen (Experiment 3) or if the listener was actively involved in the task (Experiment 1), whereas this was not necessary for salient reduction types. Interestingly, the generalization effects across reduction types varied in strength across experiments, which suggests that at least some part of learning is susceptible to our experimental manipulations. Attention as measured by task involvement (Experiment 1) seems to be of greater importance than predictability (Experiment 2) in dealing with reductions. However, the combination of these two factors (Experiment 3) yielded only a trend in the expected direction.

Moreover, the time-course results suggest that the point in time when learning about reductions takes effect may depend on the specificity of the learning process. Facilitative and inhibitory generalization effects across reduction types, which are likely not specific to any segments or words in our study, were observed early in the fixation data throughout the study (from 200 to 300 ms after target onset respectively). The inhibitory effect in Experiment 3 also emerged early (around 500 ms after target onset) but reached marginal significance only late (at 1200 ms). In contrast, the effect for generalization within reduction type in Experiment 3 was quite late (starting at 1100 ms after target onset). The word-specific effect found in Experiment 1 was equally late. The former may be explained with the kind of mapping procedure participants have to apply. Listeners learned that this particular speaker was likely to pronounce a /b/ as an [m] and hence that an existing sound ([m]) mapped onto two categories for that speaker (/m/ and /b/). Their perception of

an [m] might therefore have shifted from judging it as /m/ in most cases to judging it as /m/ in 80% and as /b/ in 20% of the cases. With this kind of learning, an initial signal-driven hypothesis strongly favors the canonical form, and only when later-arriving segments rule that form out can the learning take effect. Therefore, as soon as listeners receive evidence that a particular sound can map onto more than one category, the rule-based learning process likely needs more time to take effect. Similar reasoning can be applied to word-specific learning. At some point in time, the activation of *Parijs* "Paris" has to win over the activation of *prijs* "price" when hearing the reduced speech input *P'rijs*. Initially, the activation of *prijs* is likely to be stronger as this meaning is encountered much more frequently. Speaker-specific information (e.g., on the tendency of this speaker to reduced words like *Parijs*) then has to kick in and shift the weights in favor of the candidate *Parijs*. This may not happen immediately.

As stated before, all measures in these experiments (RT, accuracy, eye movements) could reflect improvements in spoken-word recognition due to adaptation to deviant pronunciations. However, as we did not push participants to click as fast as possible, it is perhaps not surprising that RTs did not show adaptation effects in any of the experiments. The eye-tracking data may be the more sensitive measure of adaptation because the fixation behavior does not necessarily entail conscious decision processes (unlike the click responses). Note that although we found a word-specific learning effect for vowel-deleted words in Experiment 2 in the accuracy data but not in the eye-tracking data, the eye-tracking data did show a non-significant trend in the expected direction. Note also that there were fewer participants in Experiment 2 than in Experiment 1 (60 vs. 75). It is thus possible that with more participants both measures might have shown significant effects.

Finally, it has to be noted that the learning effects (i.e., the differences between groups) were rather subtle. As stated at the beginning of the introduction, we investigated *whether* and *how* adaptation plays a role in the recognition of reduced forms. As discussed, the small learning effects speak for both episodic storage and abstraction in response to different challenges posed by different forms of reduction (answering the *how* question). Additionally, the group differences were consistently small, despite a reasonably large N (at least 60 participants in each experiment). Adaptation effects of considerable magnitude have been found with much smaller groups in conceptually similar experiments (e.g., Reinisch et al., 2013). This seems to indicate that short-term adaptation is only one piece of the puzzle concerning how we are able to understand speech despite considerable phonological reduction (answering the *whether* question).

## CONCLUSION

The present study provided evidence that listeners use a wide variety of adaptation mechanisms when dealing with reduced forms. Word-specific learning effects showed that reduced forms are sometimes stored as such in the mental lexicon. If possible, that is, if the input was sufficiently consistent, abstraction rules were generated based on the reduced speech input and applied to new reduced words. In the setting of the present study, this was only



successful for new words of the same reduction type. If the input was too inconsistent, listeners showed perceptual flexibility and were able to deal with various reduction types. The interplay of abstraction processes and perceptual adjustments may come at a cost if abstract mapping rules are already in place. The perceptual system might then not be flexible enough to allow rapid accommodation to inconsistent reductions. To conclude, both episodic and abstractionist modes of lexical access, as well as perceptual flexibility, play a role in recognizing reduced word forms.

## AUTHOR CONTRIBUTIONS

This work is part of Katja Poellmann's Ph. D. project.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00437/abstract>

## ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) under the priority program 1234 "Phonological and phonetic competence: Between grammar, signal processing, and neural activity." Some of these data were presented at the Nijmegen workshop on "Production and Comprehension of Conversational Speech" and at the 13th NVP Winter Conference in Egmond aan Zee, The Netherlands, in December 2011.

## REFERENCES

- Boersma, P., and Weenink, D. (2010). *Praat: Doing Phonetics by Computer (Version 5.1.26)*. Amsterdam: Institute of Phonetic Sciences.
- Brouwer, S., Mitterer, H., and Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Lang. Cogn. Process.* 27, 539–571. doi: 10.1080/01690965.2011.555268
- Brouwer, S., Mitterer, H., and Huettig, F. (2013). Discourse context and the recognition of reduced and canonical spoken words. *Appl. Psycholinguist.* 34, 519–539. doi: 10.1017/S0142176411000853
- Clarke, C. M., and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.* 116, 3647–3658. doi: 10.1121/1.1815131
- Cutler, A., Eisner, F., McQueen, J. M., and Norris, D. (2010). "How abstract phonemic categories are necessary for coping with speaker-related variation," in *Laboratory Phonology 10*, eds C. Fougere, B. Kühnert, M. D'Imperio, and N. Vallée (Berlin: de Gruyter), 91–111.
- Dixon, P. (2008). Models of accuracy in repeated-measures designs. *J. Mem. Lang.* 59, 447–456. doi: 10.1016/j.jml.2007.11.004
- Ernestus, M. (2000). *Voice Assimilation and Segment Reduction in Casual Dutch. A Corpus-based Study of the Phonology-Phonetics Interface*. Ph.D. thesis, LOT, Utrecht.
- Ernestus, M., Baayen, H., and Schreuder, R. (2002). The recognition of reduced word forms. *Brain Lang.* 81, 162–173. doi: 10.1006/brln.2001.2514
- Ernestus, M., and Warner, N. (2011). An introduction to reduced pronunciation variants. *J. Phon.* 39, 253–260. doi: 10.1016/S0095-4470(11)00055-6
- Floccia, C., Goslin, J., Girard, F., and Konopczynski, G. (2006). Does a regional accent perturb speech processing? *J. Exp. Psychol.* 32, 1276–1293. doi: 10.1037/0096-1523.32.5.1276
- Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory [Review]. *J. Exp. Psychol.* 22, 1166–1183. doi: 10.1037/0278-7393.22.5.1166
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279. doi: 10.1037/0033-295X.105.2.251
- Hanique, I., Ernestus, M., and Schuppler, B. (2013). Informal speech processes can be categorical in nature, even if they affect many different words. *J. Acoust. Soc. Am.* 133, 1644–1655. doi: 10.1121/1.4790352
- Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007
- Jesse, A., and McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychon. Bull. Rev.* 18, 943–950. doi: 10.3758/s13423-011-0129-2
- Johnson, K. (2004). "Massive reduction in conversational American English," in *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, eds K. Yoneyama and K. Maekawa (Tokyo: The National International Institute for Japanese Language), 29–54.
- Kemps, R., Ernestus, M., Schreuder, R., and Baayen, H. (2004). Processing reduced word forms: the suffix restoration effect. *Brain Lang.* 90, 117–127. doi: 10.1016/S0093-934X(03)00425-5
- Keuleers, E., Brysbaert, M., and New, B. (2010). SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behav. Res. Methods* 42, 643–650. doi: 10.3758/BRM.42.3.643
- Manuel, S. Y. (1992). "Recovery of "deleted" schwa," in *Perilus: Papers from the Symposium on Current Phonetic Research Paradigms for Speech Motor Control*, Vol. XIV, eds O. Engstrand and C. Kylander (Stockholm: University of Stockholm), 115–118.
- McQueen, J. M., Cutler, A., and Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cogn. Sci.* 30, 1113–1126. doi: 10.1207/s15516709cog0000\_79
- McQueen, J. M., and Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *J. Acoust. Soc. Am.* 131, 509–517. doi: 10.1121/1.3664087
- McQueen, J. M., and Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Q. J. Exp. Psychol.* 60, 661–671. doi: 10.1080/17470210601183890
- Mitterer, H., and McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE* 4:e7785. doi: 10.1371/journal.pone.0007785
- Nygaard, L. C., and Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Percept. Psychophys.* 60, 355–376. doi: 10.3758/BF03206860
- Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychol. Sci.* 5, 42–46. doi: 10.1111/j.1467-9280.1994.tb00612.x
- Oostdijk, N. (2000). The spoken dutch corpus project. *ELRA Newslett.* 5, 4–8.
- Patterson, D., LoCasto, P. C., and Connine, C. M. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica* 60, 45–69. doi: 10.1159/000070453
- Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *J. Acoust. Soc. Am.* 118, 2561–2569. doi: 10.1121/1.2011150
- Reinisch, E., Weber, A., and Mitterer, H. (2013). Listeners retune phoneme categories across languages. *J. Exp. Psychol.* 39, 75–86. doi: 10.1037/a0027979
- Schuppler, B., Ernestus, M., Scharenborg, O., and Boves, L. (2011). Acoustic reduction in conversational Dutch: a quantitative analysis based on automatically generated segmental transcriptions. *J. Phon.* 39, 96–109. doi: 10.1016/j.wocn.2010.11.006
- Witteman, M. J., Weber, A., and McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attent. Percept. Psychophys.* 75, 537–556. doi: 10.3758/s13414-012-0404-y

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 October 2013; accepted: 24 April 2014; published online: 30 May 2014.  
Citation: Poellmann K, Mitterer H and McQueen JM (2014) Use what you can: storage, abstraction processes, and perceptual adjustments help listeners recognize reduced word forms. *Front. Psychol.* 5:437. doi: 10.3389/fpsyg.2014.00437  
This article was submitted to *Language Sciences*, a section of the journal *Frontiers in Psychology*.  
Copyright © 2014 Poellmann, Mitterer and McQueen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Processing of syllable stress is functionally different from phoneme processing and does not profit from literacy acquisition

Ulrike Schild<sup>1\*</sup>, Angelika B. C. Becker<sup>2</sup> and Claudia K. Friedrich<sup>1</sup>

<sup>1</sup> Developmental Psychology, University of Tübingen, Tübingen, Germany

<sup>2</sup> Biological Psychology and Neuropsychology, University of Hamburg, Hamburg, Germany

## Edited by:

Ulrike Domahs, University of Marburg, Germany

## Reviewed by:

Mathias Scharinger, Max Plank Institute for Human Cognitive and Brain Sciences, Germany  
Eva Reinisch, Ludwig Maximilian University Munich, Germany

## \*Correspondence:

Ulrike Schild, Developmental Psychology, University of Tübingen, Schleierstraße 4, D-72076 Tübingen, Germany  
e-mail: ulrike.schild@uni-tuebingen.de

Speech is characterized by phonemes and prosody. Neurocognitive evidence supports the separate processing of each type of information. Therefore, one might suggest individual development of both pathways. In this study, we examine literacy acquisition in middle childhood. Children become aware of the phonemes in speech at that time and refine phoneme processing when they acquire an alphabetic writing system. We test whether an enhanced sensitivity to phonemes in middle childhood extends to other aspects of the speech signal, such as prosody. To investigate prosodic processing, we used stress priming. Spoken stressed and unstressed syllables (primes) preceded spoken German words with stress on the first syllable (targets). We orthogonally varied stress overlap and phoneme overlap between the primes and onsets of the targets. Lexical decisions and Event-Related Potentials (ERPs) for the targets were obtained for pre-reading preschoolers, reading pupils and adults. The behavioral and ERP results were largely comparable across all groups. The fastest responses were observed when the first syllable of the target word shared stress and phonemes with the preceding prime. ERP stress priming and ERP phoneme priming started 200 ms after the target word onset. Bilateral ERP stress priming was characterized by enhanced ERP amplitudes for stress overlap. Left-lateralized ERP phoneme priming replicates previously observed reduced ERP amplitudes for phoneme overlap. Groups differed in the strength of the behavioral phoneme priming and in the late ERP phoneme priming effect. The present results show that enhanced phonological processing in middle childhood is restricted to phonemes and does not extend to prosody. These results are indicative of two parallel processing systems for phonemes and prosody that might follow different developmental trajectories in middle childhood as a function of alphabetic literacy.

**Keywords:** spoken word recognition, lexical stress, ERPs

## INTRODUCTION

Children progressively develop sensitivity to the sound structure of oral language in middle childhood (for review see Goswami and Bryant, 1990; Ziegler and Goswami, 2005). This ability appears to be pivotal for the acquisition of alphabetic writing systems. Children with dyslexia typically have difficulty with detecting or manipulating sounds (e.g., Lyytinen et al., 2004; Ziegler and Goswami, 2005). Once acquired, literacy further shapes phonological awareness. Alphabetic readers outperform illiterate participants in metalinguistic tasks, such as phoneme deletion or phoneme substitution (e.g., Castro-Caldas et al., 1998). The question emerges if progressive refinement of phonological processing in middle childhood is restricted to phonemes or if the processing of speech in general is refined at this age.

Grapheme-to-phoneme correspondence in alphabetic writing systems has been shown to modulate spoken word recognition. Alphabetic readers recognize spoken words more slowly when the words' phonemes can be spelled in different ways than when there is only one spelling for the words' phonemes (Ziegler and

Ferrand, 1998). Facilitated word recognition for words with consistent orthography is already evident when normally developing children start reading and writing (Goswami et al., 2005; Ventura et al., 2007, 2008) but is reduced or even absent for children with dyslexia (Zecker, 1991; Desroches et al., 2010). Furthermore, native language orthography appears to have an impact on the processing of non-native language (Mitterer and McQueen, 2009; Escudero and Wanrooij, 2010). Together the findings are captured by the assumption of bi-directional activating links along the pathway of representing and processing spoken language, on the one hand, and written language, on the other (e.g., Grainger and Ferrand, 1996; Grainger and Holcomb, 2009).

Evidence that the development of phonological processing in middle childhood is intimately related to alphabetic literacy comes from functional neuroimaging. By means of fMRI, Brennan et al. (2013) compared neural activation in Chinese and English 8-to-12-year-olds while performing an auditory rhyming task. Rhyming words either were consistent in orthography (e.g., pint-mint) or inconsistent in orthography (e.g., jazz-has).

Increased activation of a left-hemispheric phonological network with increasing age, enhanced activation for consistent compared to inconsistent words and a positive correlation between reading skills and superior temporal gyrus activation were found for native English children, but not for native Chinese children. The authors argue that improved phonological awareness and refined phonological processing in English speakers is related to the relatively systematic grapheme-to-phoneme correspondence in English, which contrasts to the relatively arbitrary mapping of written characters to spoken syllables in Chinese.

In line with the assumption of progressively refined phoneme processing as a function of literacy acquisition, we recently found that readers and pre-readers differ in how detailed they process sub-phonemic information in speech recognition (Schild et al., 2011). We tested pre-reading preschoolers, reading preschoolers and second graders by means of the lexical decision latencies and event-related potentials (ERPs) recorded in word onset priming. Spoken syllables (primes) were followed by spoken words (targets). The amount of phoneme overlap between primes and targets was manipulated. For all reading children and for reading adults (Friedrich et al., 2009), a condition in which primes and targets were identical (e.g., in the prime-target pair *mon-Monster* “monster”) differed from a condition in which the onset phoneme of the primes varied in one feature, namely the place of articulation, from the targets (e.g., *non-Monster*). By contrast, “Monster” was primed equally well by both primes “mon” and “non” in pre-reading children. We concluded that readers use more phoneme-relevant detail in lexical access than pre-reading children.

Phonemes are not the only type of information that spoken language entails. Prosody is another source. To establish word prosody, a speaker gives relative emphasis to a certain syllable via enhanced duration, pitch and amplitude. Therewith, phonemically identical syllables might be realized with or without stress. For example, the first syllable of the English word *music* is relatively longer, louder and has higher pitch than the first syllable of the English word *museum*. Similar to written English, written German does usually not code for syllable stress. For example, the stress difference between *August* with stress on the first syllable in spoken German (referring to a male name), and *August* with stress on the second syllable (referring to the month “August”), is not coded in the written forms of those words. For illustration purpose only, we will indicate the stressed syllable of example words by capital letters in the following article (e.g., *MU*sic and *mu*SEum, or *AU*gust and *au*GUST).

From a neurocognitive perspective, it appears that the acoustic input is decomposed into phonemes and prosody. Rapidly varying phoneme-relevant information, on the one hand, and more slowly varying prosodic information, on the other hand, are processed by different neuronal networks in adults (Zatorre and Belin, 2001; Boemio et al., 2005; Giraud et al., 2007; Giraud and Poeppel, 2012; Luo and Poeppel, 2012) and in infants (Telkemeyer et al., 2009). In line with this, Event Related Potentials (ERPs) recorded in a previous cross-modal auditory-visual priming study with adults revealed the independent processing of phonemes and pitch contours, as indicated by separate ERP phoneme priming and ERP stress priming (Friedrich et al., 2004).

Previous behavioral priming results show that adults rapidly integrate syllable stress and phonemes in ongoing speech recognition. In cross-modal auditory-visual priming, adults recognize printed words faster when they are preceded by a spoken stress matching syllable, such as the printed word *music* preceded by the spoken stressed syllable *MUS*-, than when they are preceded by a spoken stress mismatching syllable, such as *music* preceded by the spoken unstressed syllable *mus*- (see Cooper et al., 2002 for English; Soto-Faraco et al., 2001 for Spanish; and van Donselaar et al., 2005 for Dutch). Similarly, adults’ eye movements are rapidly biased by syllable stress in the visual world paradigm. For example, already before the end of the first syllable of the Dutch word *OCtopus* is encountered, Dutch participants fixate the printed version of *octopus* more frequently than they fixate the printed version of the stress competitor *okTOber* (Reinisch et al., 2010).

In the present study, we focus on the processing of syllable stress in middle childhood. Given the developing phoneme awareness in preschoolers (for review see Goswami and Bryant, 1990; Ziegler and Goswami, 2005) and the refined phoneme processing in beginning readers (Schild et al., 2011), the question emerges whether the processing of all aspects of the speech signal is shaped in middle childhood or whether the refinement of phoneme processing is a function of the acquisition of an alphabetic writing system.

Similar to our previous priming study on the processing of syllable prosody in adults (Friedrich et al., 2004), we orthogonally varied stress-overlap and phoneme-overlap between primes and targets in the present experiment. To make the paradigm appropriate for testing pre-reading children and beginning readers, we had to use a unimodal auditory design in which spoken stressed and unstressed syllables (primes) were followed by spoken disyllabic initially stressed words (targets). This resulted in four prime-target combinations: (i) Stress overlap and phoneme overlap between the prime syllable and the onset of the target word, as in the prime-target pair *MON-MONster*, (“stress-match, phoneme-match”); (ii) Pure stress overlap between the prime syllable and the onset of the target word, as in the prime-target pair *TEP-MONster* (“stress-match, phoneme-mismatch”); (iii) Pure phoneme overlap between the prime syllable and the onset of the target word, as in the prime-target pair *mon-MONster* (“stress-mismatch, phoneme-match”); or (iv) Neither stress nor phoneme overlap between the prime syllable and the onset of the target word, as in the prime-target pair *tep-MONster* (“stress-mismatch, phoneme-mismatch”).

Although unimodal auditory priming has proven to elicit earlier phoneme priming effects than cross-modal priming, other characteristic ERP deflections are largely comparable between both types of paradigms (Friedrich et al., 2009; Schild et al., 2012). Regarding the effects of different onsets of ERPs on unimodal and cross-modal priming, we concluded in our previous studies that phonological processing in the auditory modality is reflected in left-lateralized ERP differences in an early time window, ranging between 100 and 300 ms after the onset of the spoken target word (auditory N100) in adults (Friedrich et al., 2009; Schild et al., 2012) and in infants (Becker et al., 2014; but see Schild et al., 2011 for no effect in children). A left-anterior ERP phoneme priming

effect between 300 and 400 ms in both uni- and cross-modal priming, the P350 effect, has been related to matching processes between speech input and lexical representation in adults (e.g., Friedrich, 2005; Friedrich et al., 2009; Schild et al., 2011) and in children (Schild et al., 2012). Finally, an N400-like central negativity starting earlier in unimodal- than in cross-modal priming has been related to predictive phonological processing in adults (e.g., Friedrich, 2005; Friedrich et al., 2008; Schild et al., 2012), in children (Schild et al., 2011) and in infants (Becker et al., 2014). In line with the neurocognitive evidence for independent processing of phoneme-relevant and stress-relevant information (e.g., Boemio et al., 2005; Giraud et al., 2007; Telkemeyer et al., 2009) and based on our previous results (Friedrich et al., 2004), we expect to find independent ERP phoneme priming and ERP stress priming in the present study.

Comparing the processing of syllable stress in pre-readers, beginning readers and adults will enable us to draw conclusions on the middle childhood development of phonological processing. To the best of our knowledge, this study is the first to follow the development of processing of syllable stress over the time related to literacy acquisition in middle childhood. Three possible outcomes could provide insights into language development at that age. First, if the processing of the speech signal in general is refined in readers, they should use syllable stress more effectively than pre-readers. Second, if the processing of speech is refined for those aspects of the speech signal that are relevant in the alphabetic writing system, there might be no difference in how efficiently readers and pre-readers use syllable stress. Third and finally, if literacy draws processing resources away from those aspects of the speech signal that are not coded in the writing system, pre-readers might use syllable stress more efficiently than readers.

## METHODS

### PARTICIPANTS

A total of 23 pre-reading preschoolers, 24 beginning readers and 22 adults entered the analysis. Five additional participants were tested but were not included in the final analysis. Two preschoolers did not finish the experiment; and for two beginning readers and for one adult, too few EEG segments remained after artifact correction. Participant characteristics and the results of psychometric tests are summarized in **Table 1**. All children had normal or above normal IQ scores, as measured with the Raven Colored Progressive Matrices (CPM, Bulheller and Häcker, 2002). In this way we ensured that the differences between groups could not be due to general intelligence. The BISC test (Bielefelder Screening zur Früherkennung von Lese-Rechtsschreibschwierigkeiten, Jansen et al., 2002) indicated that no child was at risk for developing reading or writing impairments. Pre-reading preschoolers were not yet able to read or write words beyond their own name. Beginning readers were at the end of their second year of school. They were able to read at age-appropriate level, as confirmed by a reading test (ELFE1-6, Lenhard and Schneider, 2006). All participants were native speakers of German and were right-handed as assessed by the Edinburgh inventory (Oldfield, 1971). None of the participants reported hearing or neurological problems.

**Table 1 | Sample size (number of girls/boys and females/males, respectively), age (mean year/month for children and mean years for adults, with respective ranges), mean IQ-score (percentile rank with standard error of mean) accessed with CPM (Bulheller and Häcker, 2002) and handedness (lateralization quotient, LQ, with standard error of mean) accessed by the Oldfield Handedness Questionnaire (Oldfield, 1971) are given.**

	Sample size	Age	CPM	LQ
Pre-reading preschoolers	23 (12/11)	6.3 (5.8–6.9)	61.35 (5.12)	87.65 (2.38)
Beginning readers	24 (11/13)	7.11 (7.2–8.11)	66.17 (4.85)	89.83 (2.38)
Adults	22 (5/17)	25 (19–41)	–	85.36 (2.78)

*Pre-reading preschoolers and beginning readers showed no significant differences in CPM or LQ.*

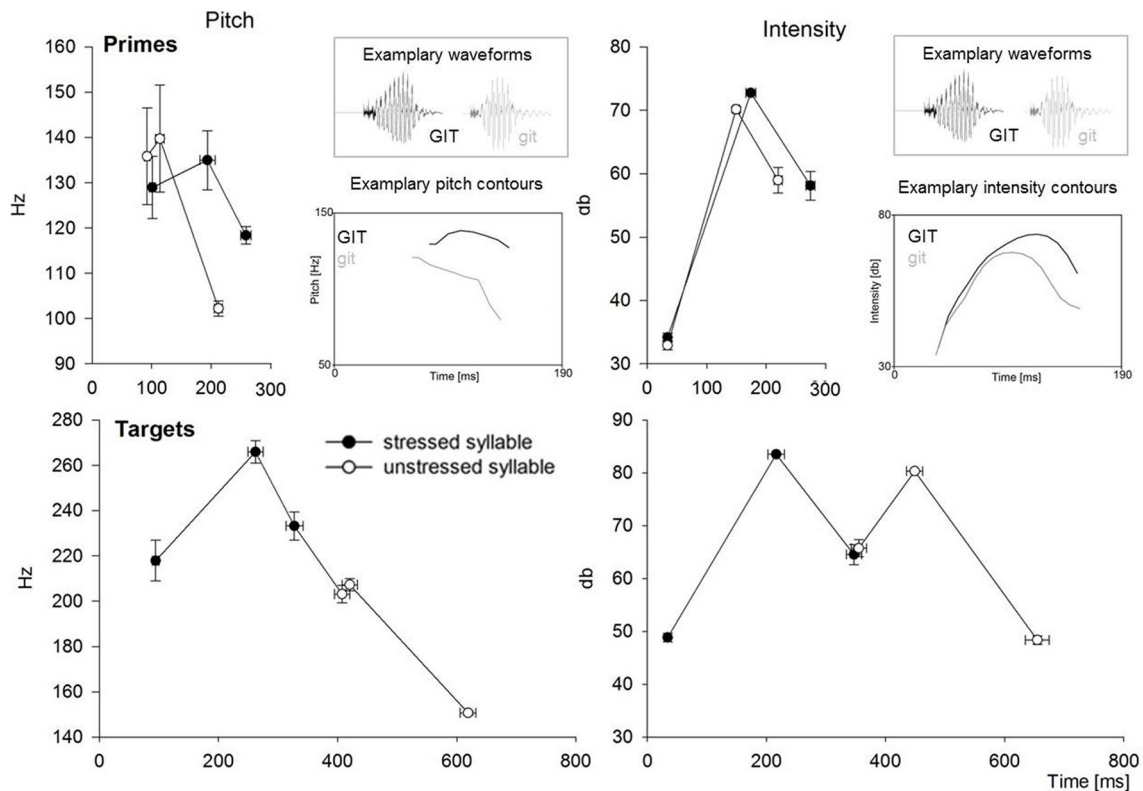
Children were recruited from local schools in Hamburg. Adults were mostly students from the University of Hamburg. They were recruited via mailing lists and internet advertisement. The children and their parents, as well as the adult participants, gave informed consent prior to their inclusion in the study. Children received a gift for their participation (child book or game). The prize of the gift matched the financial compensation of the adult participants. Adults received credit points (students of Psychology) or 8 Euros per hour as compensation for their participation in the study. The study was approved by the Ethics Committee of the German Psychological Association (Deutsche Gesellschaft für Psychologie, DGPs, 10.2006).

### MATERIALS

Forty-five monomorphemic, initially stressed disyllabic German nouns served as stimuli (see Supplementary Material). All of the words had been used in a former study in which we ensured that the words were known by young children (Schild et al., 2011). Pseudowords were created by changing the last phoneme/s of each word (e.g., *Monster* ≥ *\*Monste*).

For the primes, a male native speaker of German (a professional actor) produced the target words once with correctly applied stress (e.g., *MONSter*) and once with incorrectly applied stress (e.g., *monsTER*). We extracted the first syllable of both versions, respectively. Stressed primes were extracted from the correctly stressed version. Unstressed primes were extracted from the incorrectly stressed version. Unstressed primes were realized with full vowels because vowel reduction is not only realized via prosodic parameters but also via the phoneme-relevant parameter vowel quality. In all audio files, the onset of the stimulus was preceded by a 50 ms silent period. The cut-off for the rhymes was the end of the first syllable. If the syllable boundary spanned a plosive speech sound (e.g., *MAT-te*), the prime was cut after the closure, directly before the release.

**Figure 1** illustrates the realization of syllable stress for the primes (spoken first syllable) and the targets (spoken disyllabic word with initial stress). Amplitude and pitch measures were obtained by analyzing the whole time window of the prime syllables, of the initial syllables of the targets and of the second syllables



**FIGURE 1 |** The figure illustrates the pitch and intensity for the monosyllabic stressed and unstressed primes (above) and the disyllabic initially stressed target words (below) that were presented in the experiment. Simplified pitch and intensity contours are sketched by the mean first value, the mean maximum value and the mean last value for the monosyllabic primes, as well as for each syllable of the target words. The averaged values are given at the averaged time point they were identified in the signals for stressed and unstressed syllables respectively. Pitch and intensity values were obtained by considering the

whole syllable, because the stressed and unstressed syllables were segmentally identical and, therefore, voiced vs. unvoiced segments equally contributed to the pitch contours in both types of syllables. Error-bars indicate standard errors. Measures for stressed syllables are illustrated by black circles. Measures for unstressed syllables are illustrated by white circles. Exemplary intensity and pitch contours for the stressed prime (*GIT* taken from *GITter*, Engl. grid) and the unstressed prime (*git* taken from *\*gitTER*) illustrate the most typical contours. Waveforms of both primes are given for further illustration.

of the targets, using the software package PRAAT 5.3.17 (Boersma and Weenink, 2014). As is typical, stressed syllables were on average longer and louder than unstressed syllables. Furthermore, stressed syllables showed a pronounced longer period of rising pitch compared to unstressed syllables. This means that the maximum pitch value was reached earlier in unstressed than in stressed syllables. By contrast, the maximum intensity was reached at approximately the same time for stressed and unstressed primes. Therewith, differences in the pitch contours between the stressed and unstressed syllables appear to be earlier available in the signal than differences in the intensity contours.

Targets (words and pseudowords) were spoken by a female native speaker of German (also a professional actor). Digital audio files for each single target were extracted from those utterances. In all audio files, the onset of the stimulus was preceded by a 50 ms silent period. The same target word was presented in four different types of prime-target pairs: (i) Stress overlap and phoneme overlap between prime and target (S+P+, e.g., *MON-MONster*); (ii) stress overlap without phoneme overlap (S+P–; e.g., *TEP-MONster*); (iii) phoneme overlap without stress overlap (S–P+;

e.g., *mon-MONster*); and (iv) neither phoneme nor stress overlap (S–P–, e.g., *tep-MONster*). Thus, the stress and phonemes were manipulated independently. The same mapping was applied for pseudowords. To make the task appropriate for children, we had to adapt the lexical decision task, which contained 50% pseudowords, to a go/no-go task, which had only 25% pseudowords. Our pilot testing confirmed that the experiment would have been too long for preschoolers if we had included more pseudoword trials. Moreover, in many priming studies, a lexical decision task is used, in which participants respond to a word with one button and to a pseudoword with another button. Again, our pilot studies showed that these two response alternatives are too demanding for pre-schoolers. Therefore, we decided to use a go/no-go task with a low percentage of non-words (25%) and a single response alternative (“word”).

## DESIGN AND PROCEDURE

Each participant completed a total of 240 trials (180 target words, 60 target pseudowords). In twelve consecutive blocks, 20 trials were presented each time. Within blocks 1–3, 4–6, 7–9, and



10–12, no repetition of a target word or a pseudoword occurred. Within and across blocks, the order of trials was randomized. In sum, each participant received the same target word four times with four different pairings of primes.

Participants were comfortably seated in an electrically shielded and sound-attenuated booth. Each experimental trial started with the presentation of a “fixation smiley” (size:  $1 \times 1$  cm) at the center of a computer screen in front of the participants (distance: 70 cm). Participants were instructed to fixate on this smiley whenever it appeared. The first audio fragment (prime) was presented via loudspeakers 500 ms after the onset of the fixation smiley. The target was delivered 250 ms after offset of the fragment. The inter-stimulus interval includes the 50 ms silence from the beginning of the way file for the target. Participants were instructed to respond as quickly and accurately as possible to words but to refrain from responding when the target was a pseudoword (go/no-go task). If an overt response was given, visual feedback (size:  $3 \times 7$  cm) appeared for 2 s. A smiley different from the “fixation smiley” was presented if the participant responded correctly to a word, whereas a ghost was presented if the participant responded to a pseudoword incorrectly. If no response occurred, no feedback was delivered, and the fixation smiley remained for 3.5 s. The next trial started after a 1.5 s inter-trial interval. The loudspeakers were placed on the left and right sides of the screen. Half of the participants pressed the response button with their left index finger, and half, with their right index finger. Auditory stimuli were presented at comfortable listening sound levels of approximately 70 db. Stimulus presentation was controlled by Presentation® software (Version 14.9, Neurobehavioral Systems, Berkeley, CA, U.S.A.).

### EEG-RECORDING AND ANALYSIS

The continuous EEG was recorded at a 500 Hz sampling rate (bandpass filter 0.01–100 Hz, BrainAmp Standard, Brain Products, Gilching, Germany) from 46 active Ag/AgCl electrodes mounted in an elastic cap (Electro Cap International, Inc.) according to the international 10–20 system (two additional electrodes below the eyes, ground at position AF3). For adults, we recorded from 73 electrodes. After recording with a nose electrode reference, the continuous EEG was off-line re-referenced to an average reference and highpass-filtered by 0.3 Hz.

Eye artifacts were corrected using surrogate Multiple Source Eye Correction (MSEC) by Berg and Scherg (1994), as implemented in the Besa Research-Software® (Version 5.3, MEGIS Software GmbH; Gräfelfing, Germany). Here, brain activity is modeled by a fixed dipole model (the “surrogate model”), and spatial artifact topographies are used to correct the artifacts in the ERP data. To adjust typical artifact topographies to the individual artifact topographies, calibration trials for blinks, vertical and horizontal eye movements were recorded prior to the experiment from the children. The continuous EEG was then corrected for those eye movements by means of a principal component analysis (for details see Berg and Scherg, 1994). Because adults barely moved their eyes in the experiment, for them, only blinks out of the experiment were used and corrected. The remaining artifacts, such as slow drift or movement artifacts, were eliminated according to visual inspection. Individual electrodes showing artifacts that were not reflected in the remaining electrodes in more than

two trials were interpolated for all trials. This practice resulted in approximately 2 interpolated electrodes per participant (mean = 2.3, Standard Error of mean [SE] = 0.2; not significantly different between groups, all  $t < 1.8$ , ns).

ERP segments were computed for the target words with correct responses, starting from the beginning of the speech signal up to 1000 ms post-onset of the stimulus and having a 200 ms prestimulus baseline. All data sets included at least 19 segments in each condition (mean/SE across groups: S+P+: 35.2/0.8; S+P–: 35.4/0.7; S–P+: 36.0/0.8; S–P–: 35.2/0.8). There were no significant differences in the numbers of segments in each condition.

### DATA ANALYSIS

As in our previous study (Schild et al., 2011), responses shorter than 200 ms and longer than 2000 ms, which is approximately in the 2-standard-deviation margin, were removed from the behavioral analyses. Reaction times calculated from the onset of the words up to the participants' responses were subjected to a two-way repeated measures ANOVA with the within-participant two-level factor *Stress Overlap* (prime and target onset match vs. mismatch in stress) and *Phoneme Overlap* (prime and target onset match vs. mismatch in phonemes) and the between-participant three-level factor *Group*.

Because the ERP variance for processing different words is high, targets usually are presented several times in ERP studies so that they are heard in all possible prime-target combinations by a single participant. Consequently, target words were repeated four times in the present experiment. This procedure diverges from classical psycholinguistic designs, in which target repetitions within participants are avoided. To compare the present behavioral results with those of former studies using the classical procedure without target word repetition (Soto-Faraco et al., 2001; Cooper et al., 2002 for Spanish, van Donselaar et al., 2005), we analyzed the first presentation of each target word in addition to the analysis of all presentations.

To analyze the ERP effects, two additional factors were used, *Hemisphere* (left vs. right electrode sites) and *Region* (anterior vs. posterior electrode sites). We calculated the same ROIs as in our former study, namely four lateral ROIs (anterior left: F9, F7, F3, FT9, FT7, FC5, FC1, T7, C5; anterior right: F10, F8, F4, FT10, FT8, FC6, FC2, T8, C6; posterior left: C3, TP9, TP7, CP5, CP1, P7, P3, PO9, O1; posterior right: C4, TP10, TP8, CP6, CP2, P8, P4, PO10, O2) and two central ROIs (anterior: FPz, AFz, Fz, FCz; posterior: Cz, Pz, POz, Iz). In case of significant interactions,  $t$ -tests were computed to evaluate the differences among conditions. ERP analysis was based on average references. For ERP analysis, only interactions including the factor *Stress Overlap*, the factor *Phoneme Overlap* or both factors are reported. Data analysis was performed with SPSS® software (Version 19, IBM®).

### RESULTS

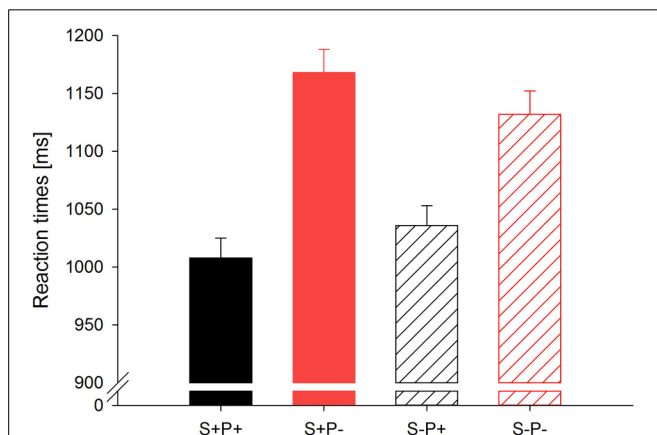
The mean reaction times for each group and conditions for the first presentation and overall are given in Table 2, and illustrated for the first presentation in Figure 2.



**Table 2 | Mean reaction times in milliseconds (and standard error of mean) are shown for each group and each condition**

	First 60 trials (no target repetition)				All trials (four target repetitions)			
	S+P+	S+P–	S–P+	S–P–	S+P+	S+P–	S–P+	S–P–
Pre-reading children	1093 (29)	1295 (25)	1149 (30)	1275 (33)	1082 (27)	1199 (23)	1104 (23)	1176 (27)
Beginning readers	1049 (25)	1228 (27)	1068 (22)	1152 (25)	1083 (23)	1177 (25)	1083 (22)	1154 (27)
Adults	904 (24)	1022 (21)	944 (21)	986 (20)	911 (22)	943 (21)	916 (20)	931 (19)
<b>Combined groups</b>	<b>1018 (18)</b>	<b>1185 (20)</b>	<b>1056 (17)</b>	<b>1140 (21)</b>	<b>1028 (17)</b>	<b>1110 (19)</b>	<b>1037 (16)</b>	<b>1090 (19)</b>

The results for the first target presentation (without target repetition) are shown in the left columns. The results for all trials (with four target repetitions) are shown in the right columns. Abbreviations for the four conditions are as follows: “S+P+” for stress match, phoneme match (e.g., MON–MONster); “S+P–” for stress match, phoneme mismatch (e.g., TEP–MONster); “S–P+” for stress mismatch, phoneme match (e.g., mon–MONster); and “S–P–” for stress mismatch, phoneme mismatch (e.g., tep–MONster).



**FIGURE 2 | Mean reaction times across all groups for each condition for the first 60 trials (trials without repetition of the target word).** Error bars indicate standard errors. The abbreviations of the four conditions are as follows: “S+P+” for stress match, phoneme match (e.g., MON–MONster); “S+P–” for stress match, phoneme mismatch (e.g., TEP–MONster); “S–P+” for stress mismatch, phoneme match (e.g., mon–MONster); and “S–P–” for stress mismatch, phoneme mismatch (e.g., tep–MONster). All conditions were significantly different from each other.

### REACTION TIMES FOR THE FIRST PRESENTATION OF THE TARGET WORDS

The ANOVA for the first presentation revealed a main effect of the factor *Group*,  $F_{(2, 66)} = 26.2$ ,  $p < 0.001$ , a main effect of the factor *Phoneme Overlap*,  $F_{(1, 66)} = 247.4$ ,  $p < 0.001$ , and an interaction between the factors *Phoneme Overlap* and *Group*,  $F_{(2, 66)} = 9.5$ ,  $p < 0.001$ . Crucially, there was an interaction between the factors *Phoneme Overlap* and *Stress Overlap*,  $F_{(1, 66)} = 33.2$ ,  $p < 0.001$ .

The main effect of the factor *Group* indicated that adults responded faster than children. The main effect of the factor *Phoneme Overlap* indicated that all participants responded faster when primes and target onsets shared phonemes than when they shared no phonemes. Follow-ups of the interaction of the factors *Phoneme Overlap* and *Group* indicated that the factor *Phoneme Overlap* was significant for each group, all  $F \geq 74.2$ ,  $p < 0.001$ . The mean difference for phoneme match and phoneme mismatch was 79 ms for the adults, 164 ms for the preschoolers and 130 ms for the second graders. Both groups of children showed

stronger phoneme priming effects than adults,  $F \geq 10.1$ ,  $p < 0.01$ . However, the groups of children did not differ significantly from each other,  $F < 2.4$ , ns.

Following up the interaction of the factors *Phoneme Overlap* and *Stress Overlap*, *post-hoc* comparisons indicated that all single conditions differed significantly from each other, all  $t_{(68)} \geq 4.3$ ,  $p \leq 0.001$ . The fastest responses were made when the prime and target onset shared stress and phonemes (S+P+), whereas slowest responses were made when the prime and target onset shared stress but differed in phonemes (S+P–).

### REACTION TIMES OVERALL (FOUR REPETITIONS OF THE TARGET WORDS)

The ANOVA over all four repetitions of the targets yielded similar results as the ANOVA for the first presentation of the target words; namely, a main effect of the factor *Group*,  $F_{(2, 66)} = 26.6$ ,  $p < 0.001$ , a main effect of the factor *Phoneme Overlap*,  $F_{(1, 66)} = 290.3$ ,  $p < 0.001$ , and an interaction of the factors *Phoneme Overlap* and *Group*,  $F_{(2, 66)} = 30.5$ ,  $p < 0.001$ , were observed. Again, there was an interaction of the factors *Phoneme Overlap* and *Stress Overlap*,  $F_{(1, 66)} = 12.2$ ,  $p < 0.01$ .

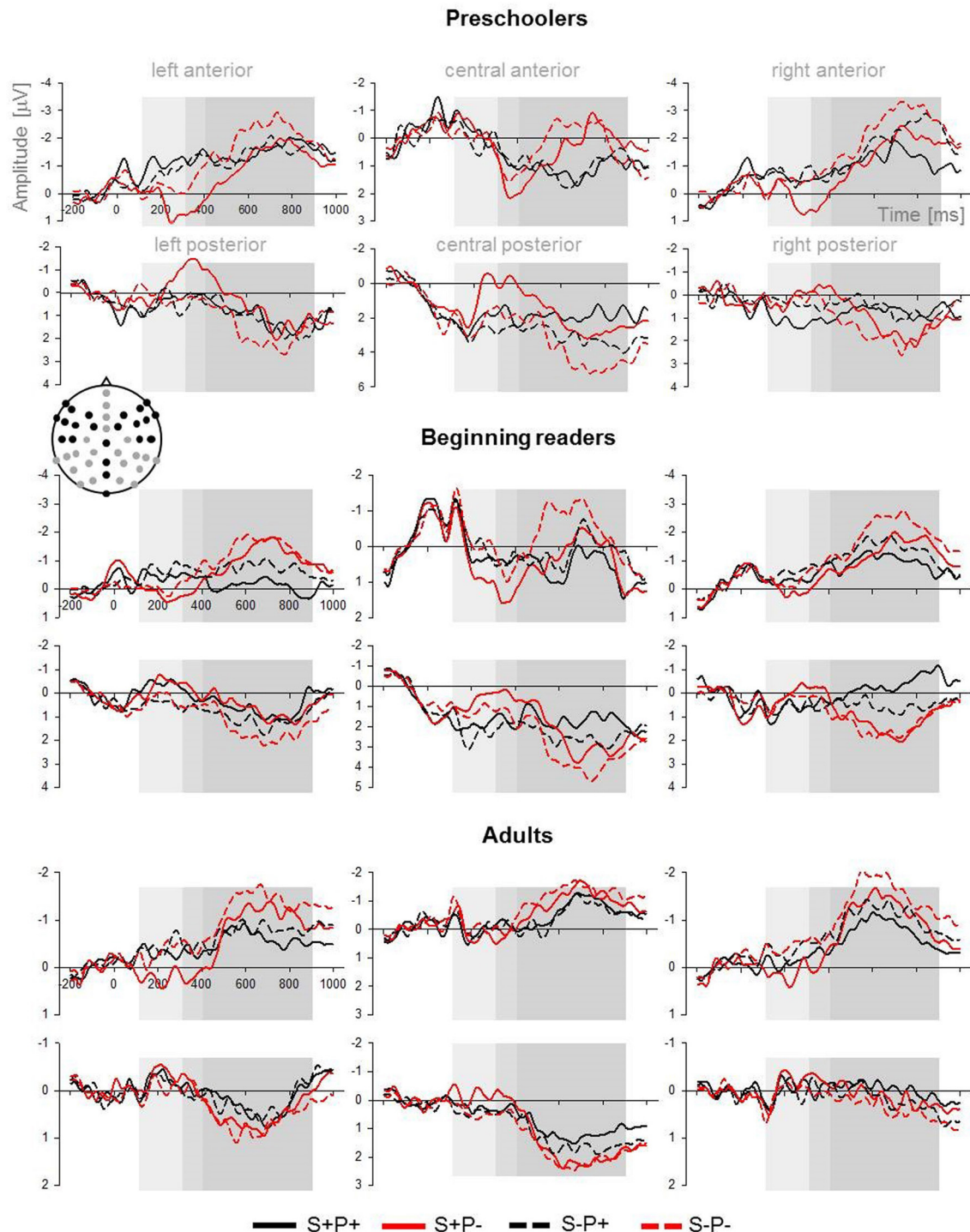
Similar to the results for the first presentation, the main effect of the factor *Group* over all blocks indicated that adults responded faster than children. The main effect of the factor *Phoneme Overlap* indicated that all participants responded faster when the primes and target onsets shared phonemes than when they shared no phonemes. Follow-ups of the interaction of the factors *Phoneme Overlap* and *Group* indicated that there was a significant phoneme priming effect for each group, all  $F \geq 26.7$ ,  $p < 0.001$ . Both groups of children showed stronger phoneme priming than adults,  $F \geq 41.0$ ,  $p < 0.001$ . The groups of children did not differ from each other,  $F < 1.3$ , n.s. The mean difference for phoneme-matching and phoneme-mismatching was 24 ms for adults, 95 ms for preschoolers and 83 ms for second graders.

Again, follow-ups of the interaction of the factors *Phoneme Overlap* and *Stress Overlap* indicated fastest responses when the prime and target onset shared stress and phonemes (S+P+). The slowest responses were obtained when the targets' first syllables shared stress but differed in the phonemes from their preceding primes (S+P–). *Post-hoc* comparisons revealed significant differences among all conditions,  $t_{(68)} \geq 3.4$ ,  $p \leq 0.001$ ,

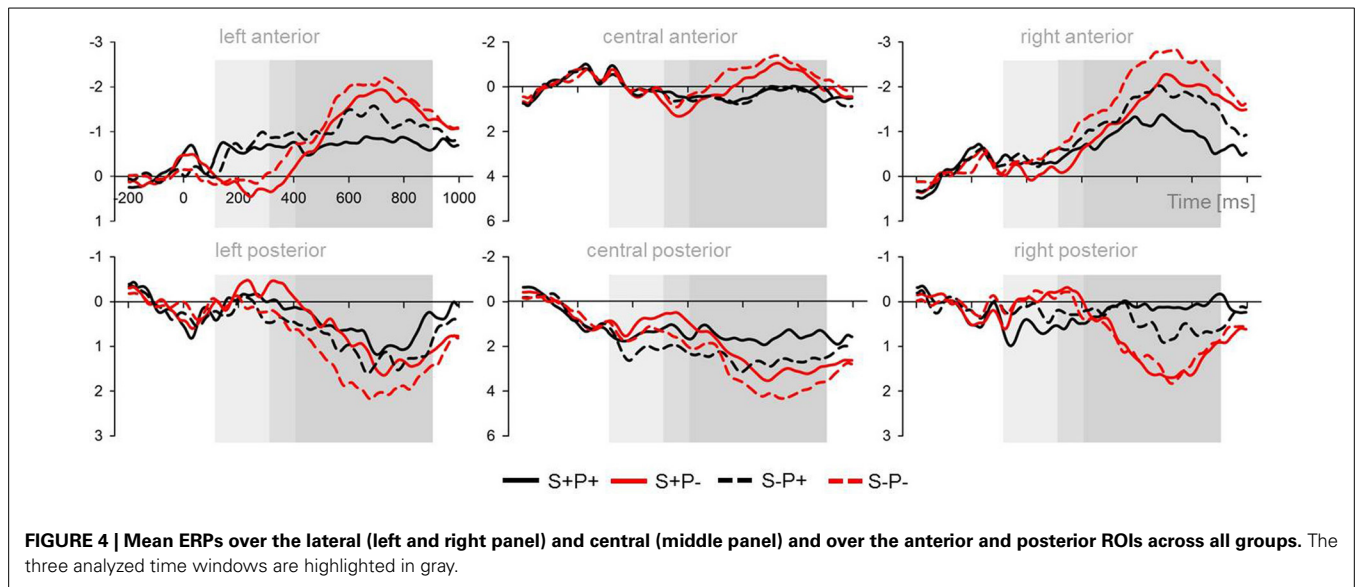
except in the case of targets that shared phonemes but did or did not diverge in stress from their preceding primes (S+P+ vs. S–P+), which was significant at the trend level only,  $t_{(68)} = 1.91$ ,  $p = 0.067$ .

### EVENT-RELATED POTENTIALS

The mean ERPs for each of the three groups are displayed in **Figure 3**. The mean ERPs across all groups for the four ROIs can be seen in **Figure 4**. We collapsed the ERP over the groups



**FIGURE 3 |** Mean ERPs over the lateral (left and right) and central and over the anterior and posterior ROIs for each group. Black and gray dots on the head montage indicate the electrode positions that contributed to the ROIs. The three time windows analyzed in greater detail are highlighted in gray.



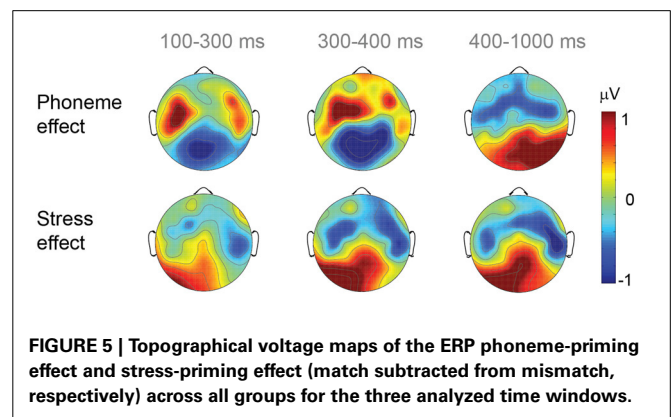
because, in the time windows from 100 to 400 ms, no group effects were observed. For topographical voltage maps of the phoneme and stress-priming effects, see **Figure 5**. According to consecutive 100-ms time window analyses (see Supplementary Material) and according to previous auditory priming studies (Friedrich et al., 2009; Schild et al., 2011, 2012), we tested the mean ERP amplitudes in three time windows in detail: (i) a time window ranging between 100 and 300 ms addressing auditory phonological processing (N100); (ii) a time window ranging between 300 and 400 ms addressing abstract lexical processing (P350) and predictive phonological processing (central negativity); and (iii) a time window ranging from 400 to 1000 ms capturing extended ERP phoneme priming and ERP stress priming.

#### Time window 100–300 ms (auditory N100)

**Lateral Electrodes.** The overall ANOVA of the lateral ROIs revealed interactions of the factor *Phoneme Overlap* with the factor *Hemisphere*,  $F_{(1, 66)} = 3.7$ ,  $p = 0.05$  and with the factor *Region*,  $F_{(1, 66)} = 8.4$ ,  $p = 0.005$ . The overall ANOVA of the lateral ROIs also revealed an interaction between the factors *Stress Overlap* and *Region*,  $F_{(1, 66)} = 4.9$ ,  $p = 0.03$ .

Follow-ups revealed main effects of the factor *Phoneme Overlap* over the left hemisphere,  $t_{(68)} = 3.4$ ,  $p = 0.001$ , and over anterior regions,  $t_{(68)} = 3.9$ ,  $p < 0.001$ . Prime-target pairs matching in phonemes elicited more negative amplitudes than prime-target pairs mismatching in phonemes. There was no significant difference between both conditions over the right hemisphere, and a trend for reversed amplitude differences between conditions over posterior regions,  $t_{(68)} = 1.9$ ,  $p = 0.06$ . Furthermore, follow-ups revealed a main effect of the factor *Stress Overlap* over posterior regions,  $t_{(68)} = 3.1$ ,  $p = 0.003$ . Amplitudes for stress match were more negative than amplitudes for stress mismatch. There was no main effect of stress over anterior regions.

**Central Electrodes.** The overall ANOVA of the central ROIs revealed an interaction between the factors *Phoneme Overlap* and



*Region*,  $F_{(1, 66)} = 4.6$ ,  $p = 0.04$ , indicating an effect for the posterior ROI that showed the same amplitude difference as was obtained for posterior lateral ROIs,  $t_{(68)} = 3.5$ ,  $p = 0.001$ .

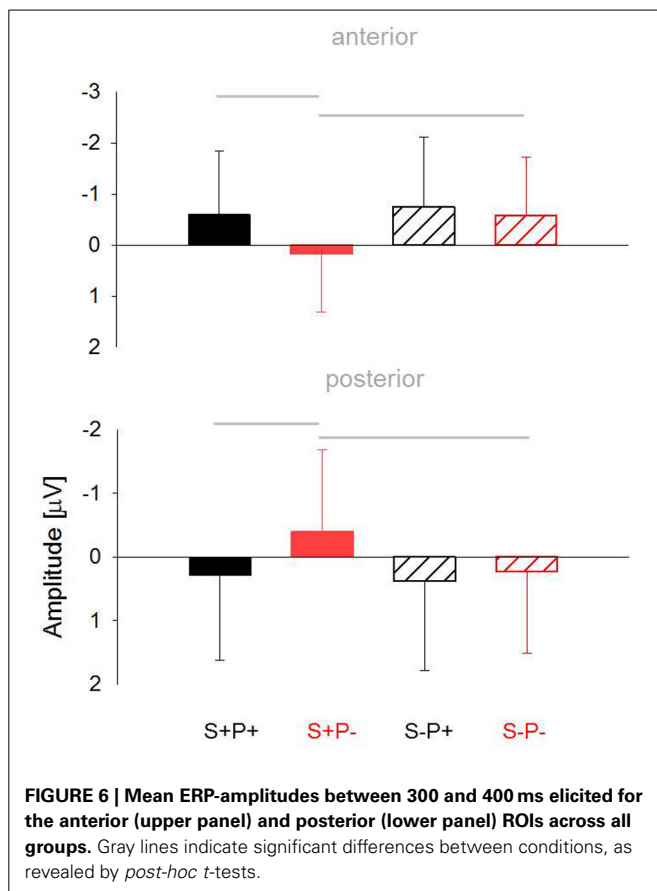
Neither over lateral ROIs nor over midline ROIs were any interactions between the factors *Stress Overlap* and *Phoneme Overlap* observed in the first time window.

#### Time window 300–400 ms (P350 and central negativity)

**Lateral Electrodes.** In this time window, we found an interaction of the factors *Phoneme Overlap*, *Stress Overlap* and *Region*,  $F_{(1, 66)} = 4.1$ ,  $p = 0.05$ , for the lateral electrodes. Follow-up analysis revealed a significant interaction between the factors *Phoneme Overlap* and *Stress Overlap* for the anterior regions,  $F_{(1, 68)} = 4.5$ ,  $p = 0.04$ , and a trend level effect for the posterior regions,  $F_{(1, 68)} = 3.4$ ,  $p = 0.07$ . Both interactions are illustrated in **Figure 6**. It appeared that the condition (S+P−) showing the slowest behavioral responses differed in ERP amplitudes from all other conditions, all  $t_{(68)} \geq 3.5$ , all  $p \leq 0.001$ . All remaining conditions did not differ from one other  $t_{(68)} < 1.1$ , ns.

**Central Electrodes.** The overall ANOVA of the central ROIs revealed an interaction between the factors *Phoneme Overlap* and





*Region*,  $F_{(1, 66)} = 24.0$ ,  $p < 0.001$ , and an interaction between the factors *Stress Overlap* and *Region*,  $F_{(1, 66)} = 17.5$ ,  $p < 0.001$ . Follow-ups of effects of the factor *Phoneme Overlap* revealed significantly more negative amplitudes for matching compared to mismatching phonemes over the anterior midline,  $t_{(68)} = 3.0$ ,  $p = 0.004$ . This pattern was reversed over the posterior midline,  $t_{(68)} = 4.4$ ,  $p < 0.001$ . Follow-ups of effects of the factor *Stress Overlap* revealed that stress-matching conditions elicited more negative amplitudes than stress-mismatching conditions over the posterior regions,  $t_{(68)} = 4.3$ ,  $p < 0.001$ .

#### Time window 400–1000 (Extended processing)

**Lateral Electrodes.** The overall ANOVA of the lateral ROIs revealed significant interactions of the factor *Phoneme Overlap* with the factor *Region*,  $F_{(1, 66)} = 23.1$ ,  $p < 0.001$ , and with the factors *Hemisphere* and *Region*,  $F_{(1, 66)} = 5.7$ ,  $p = 0.02$ . Both interactions were modulated by a four-way interaction of the factors *Hemisphere*, *Region*, *Phoneme Overlap* and *Group*,  $F_{(2, 66)} = 3.2$ ,  $p = 0.05$ . The overall ANOVA of the lateral ROIs also revealed a significant interaction of the factors *Stress Overlap* and *Region*,  $F_{(1, 66)} = 23.1$ ,  $p < 0.001$ .

Follow-up ANOVAS for each group separately revealed that only the preschoolers showed a three-way interaction of the factors *Phoneme Overlap*, *Hemisphere* and *Region*,  $F_{(1, 22)} = 7.0$ ,  $p = 0.02$ . Over right posterior regions, phoneme-matching conditions elicited more negative amplitudes than phoneme-mismatching conditions,  $t_{(22)} = 2.4$ ,  $p = 0.03$ . Both reading

groups, the beginning readers and the adults, showed interactions of *Phoneme Overlap* and *Region*, both  $F > 20.0$ ,  $p < 0.001$ . For both groups, prime-target pairs mismatching in phonemes elicited more negative amplitudes than prime-target pairs matching in phonemes over anterior regions. The reversed pattern was obtained over posterior regions, all  $t > 3.9$ ,  $p \leq 0.01$ .

Follow-ups of effects of the factor *Stress Overlap* revealed that over anterior regions, the amplitudes of the stress-mismatching conditions were more negative than the amplitudes of the stress-matching conditions,  $t_{(68)} = 5.5$ ,  $p < 0.001$ . This effect was reversed over posterior regions,  $t_{(68)} = 3.7$ ,  $p < 0.001$ .

**Central Electrodes.** The overall ANOVA of the central ROIs revealed a significant interaction of the factors *Phoneme Overlap* and *Region*,  $F_{(1, 66)} = 30.0$ ,  $p < 0.001$ , which was not modulated by the factor group. Furthermore, there was a significant interaction of the factors *Stress Overlap* and *Region*,  $F_{(1, 66)} = 21.6$ ,  $p < 0.001$ .

Follow-ups of *Phoneme Overlap* effects revealed that the amplitudes of phoneme-mismatching conditions were more negative over the anterior midline than the amplitudes of phoneme-matching conditions,  $t_{(68)} = 3.8$ ,  $p < 0.001$ . The effect was reversed over the posterior midline,  $t_{(66)} = 3.5$ ,  $p \leq 0.001$ . Follow-ups of *Stress Overlap* effects revealed the same amplitude differences as for the lateral ROIs over both the anterior midline,  $t_{(68)} = 1.86$ ,  $p = 0.07$ , and over the posterior midline,  $t_{(68)} = 4.7$ ,  $p < 0.001$ .

Neither over lateral ROIs nor over midline ROIs were any interactions between the factors *Stress Overlap* and *Phoneme Overlap* observed in the third time window.

In summary, the ERP data were quite comparable for preschoolers, beginning readers and adults. For all groups, phoneme priming started at approximately 100 ms, and stress priming started at approximately 200 ms (see Supplementary Material). Across all three larger time windows, the ERPs of all groups showed independent ERP priming effects for prime-target overlap in phonemes, on the one hand, and for prime-target overlap in stress, on the other hand. ERP phoneme priming was characterized by enhanced N100 for phoneme match and enhanced P350 and central negativity for phoneme mismatch. ERP stress priming was characterized by sustained enhanced negativity for stress match. Only in the time window ranging between 300 and 400 ms did phoneme priming and stress priming interact over lateral electrodes. Nevertheless, even in this time window, independent phoneme priming and stress priming was obtained over the midline electrodes.

## DISCUSSION

The present study focused on the processing of syllable stress in middle childhood. We tested pre-readers and beginning readers, as well as adults. Behavioral and ERP stress priming were comparable across groups. Thus, we can discard the first hypothesis stating that the processing of the speech signal in general is improved in readers, and also the third hypothesis stating that the readers withdraw processing resources from aspects of the speech signal that have no correspondence with the writing system. Instead, adults, pre-readers and alphabetic readers appeared

to similarly exploit syllable stress. Together the present results speak for the second hypothesis, stating that alphabetic readers' sensitivity is not enhanced regarding an aspect of the speech signal that does not correspond with the writing system, namely syllable stress.

The group effects in the present data suggest that refined speech processing in middle childhood is restricted to phonemes. The behavioral data indicate stronger phoneme priming effects, but not stronger stress priming effects, in children compared to adults. The ERPs point to a unique late ERP response to phoneme priming for preschoolers, but stress priming does not show a unique ERP response for any group. Together, these results reveal that, in middle childhood and especially at the preschool ages, phonological awareness might drive portions of the phoneme priming effects. That is, preschoolers and beginning readers appear to be especially sensitive to phonemes but do not modulate their processing of syllable stress. Thus, enhanced phonological processing in middle childhood appears to be restricted to those aspects of the speech signal that are relevant for acquiring an alphabetic writing system, namely phonemes, without generalizing to aspects of the speech signal that are not typically encoded in the writing system, namely prosody.

The second major finding of this study regards the independent processing of prosody and phonemes, as indicated by separate ERP phoneme priming and ERP stress priming. We uncovered that the main effects of stress overlap and the main effects of phoneme overlap did not interact in the first and third time window analyzed for the ERPs. Independent ERP phoneme priming and ERP stress priming in the same time windows provides evidence for two separate processing systems operating in parallel. This confirms the conclusion of independent processing of stress and phonemes that we have formerly drawn from ERPs recorded in cross-modal auditory-visual priming with adults (Friedrich et al., 2004).

Although ERPs allow only restricted conclusions about the localization of neuronal sources, different topographies of ERP phoneme priming and ERP stress priming support our conclusion of independent processing systems and are informative about the processing of stress. The left-lateralization of ERP phoneme priming replicates previous results obtained with unimodal auditory word onset priming (Friedrich et al., 2009; Schild et al., 2012) and cross-modal word onset priming (Friedrich et al., 2004, 2008; Friedrich, 2005). Bilateral stress priming replicates a previous result obtained with cross-modal auditory-visual word onset priming (Friedrich et al., 2004). The left-lateralization of phoneme priming is in line with the "asymmetric sampling in time" (AST) hypothesis stating that acoustic information varying on a small time-scale is processed predominantly in the left hemisphere (e.g., Poeppel, 2003; Poeppel et al., 2008). However, the AST hypothesis also states that the processing of acoustic information varying on a larger time-scale, such as syllable stress, is lateralized to the right hemisphere. This assumption is not confirmed by the present and previous bilateral ERP stress priming effects. Together our findings are in accordance with a meta-analysis of lesion literature revealing that linguistic prosodic perception is under bihemispheric control (Witteman et al., 2011).

Regarding behavioral stress priming, the present results obtained with a unimodal auditory paradigm can be integrated within previous work using a cross-modal priming paradigm. Similar to the former studies, we obtained the fastest responses for combined prime-target overlap in syllable stress and phonemes (see Soto-Faraco et al., 2001; Cooper et al., 2002; Friedrich et al., 2004; van Donselaar et al., 2005). This result reveals that pre-readers and readers rapidly integrate phonemes and prosody in ongoing spoken word recognition.

Most astonishingly, stress overlap without phoneme overlap elicited the slowest behavioral responses in the present study. This condition has been previously realized only in a single cross-modal priming study (Friedrich et al., 2004). There, behavioral responses for stress match were faster compared to stress mismatch. Here, we speculate that the enhanced response latencies for stress match in the present unimodal study result from a violation of basic rhythmic properties of speech in the stress match condition for initially stressed targets. In that condition, the stressed prime syllable is immediately followed by the stressed onset syllable of the target word. The juxtaposition of two stressed syllables, referred to as a "stress clash," violates the regularly alternating sequence of stressed and unstressed syllables in continuous speech (Liberman and Prince, 1977; Tomlinson et al., 2013). The assumption that "stress clashes" delay the processing of stress-matching targets in unimodal priming has to be validated by adding initially unstressed targets to future designs.

ERP phoneme priming, as reflected in the auditory N100, in the P350 effect and in the central negativity, was largely comparable with the results of previous studies. Previously, enhanced left-lateralized negative-going amplitudes for phoneme match compared to phoneme mismatch have been obtained for adults in the N100 time window (100 to 300 ms; Friedrich et al., 2009; Schild et al., 2012), but not for children (Schild et al., 2011). Similarly, enhanced anterior positivity for phoneme mismatch has been obtained for adults and children in the P350 time window (300 to 400 ms). The bilateral distribution of the anterior P350 effect in the present study is integrated into a heterogeneous pattern of results regarding the lateralization of this ERP deflection, for which a bilateral distribution in adults (Schild et al., 2012) and pre-readers (Schild et al., 2011) has been obtained, in addition to a left-lateralized distribution in adults (Friedrich et al., 2009) and beginning readers (Schild et al., 2011).

The topography and polarity of amplitude differences characterizing ERP stress priming differed from ERP phoneme priming. Reversed to phoneme priming, the mean ERP amplitudes for stress match were more negative than the mean ERP amplitudes for stress mismatch starting at 200 ms after the target word onset. The bilateral posterior distribution relates ERP stress priming to N400-like central negativity and therewith to predictive phonological processing in unimodal auditory priming. Enhanced negativity for stress match compared to stress mismatch reflects that stress match is somewhat unexpected. Again, the atypical sequence of two stressed syllables in both stress match conditions might be relevant here. The stressed prime syllable followed by the stressed initial syllable of the target word violates the expectation of an alternating sequence of stressed and unstressed syllables in natural speech (Liberman and Prince,



1977). Enhanced N400 amplitudes for stress clash in a sentence context have been recently reported (Bohn et al., 2013). In other words, together with the behavioral priming results, we might interpret the enhanced central negativity for stress match as reflecting an unexpected stress clash.

Only between 300 and 400 ms was there an interaction between ERP phoneme priming and ERP stress priming. This interaction effect somewhat parallels the behavioral data. The condition that elicited the slowest responses, namely stress overlap without phoneme overlap (S+P–), also diverged in the P350 effect from the other conditions. Because a similar interaction was found over the anterior and posterior regions, we cannot unambiguously relate this event to either the P350 or the central negativity. However, a unifying interpretation of the data should focus on expectancy mechanisms. It appears that the target in the condition S+P– was the least expected, as the remaining three conditions were somehow still primed. S+P+ and S–P+ are primed by phoneme overlap with their preceding primes, whereas S–P– fulfills the expected pattern of alternating syllable stress between prime syllable and target syllable. This *post-hoc* interpretation must be examined further in future research.

In conclusion, we did not find different processing of syllable stress for pre-readers and readers in the present study. This contrasts to the evidence for enhanced and refined phoneme processing in readers that we found in the present study and in a former study (Schild et al., 2011). Thus, although developmental maturation and vocabulary growth might exert an influence on phonological processing throughout childhood (Walley et al., 2003) the present and previous results might be best explained by the influence of literacy. We conclude that literacy specifically improves the processing of those aspects of speech that find correlates in the written signal. Together these results converge to the conclusion of two separate processing streams for phonemes and prosody. ERPs point to functionally and anatomically distinct networks devoted to process both types of information. Age-related differences reveal that the processing of phonemes, but not the processing of prosody is modulated by literacy acquisition.

## ACKNOWLEDGMENTS

The work was supported by a grant of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, FR 2591/1-2) awarded to Claudia K. Friedrich and Brigitte Röder and a Starting Independent Investigators Grant of the European Research Council (ERC, 209656 Neurodevelopment) awarded to Claudia K. Friedrich. We are grateful to Anne Bauch and Leon Skoba for their assistance in collecting the data.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00530/abstract>

## REFERENCES

Becker, A. B. C., Schild, U., and Friedrich, C. K. (2014). ERP correlates of word onset priming in infants and young children. *Dev. Cogn. Neurosci.* 9, 44–55. doi: 10.1016/j.dcn.2013.12.004

- Berg, P., and Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalogr. Clin. Neurophysiol.* 90, 229–241. doi: 10.1016/0013-4694(94)90094-9
- Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* 8, 389–395. doi: 10.1038/nn1409
- Boersma, P., and Weenink, D. (2014). *Praat: Doing Phonetics by Computer [Computer Program]. Version 5.3.63*. Available online at: <http://www.praat.org/> (Accessed 24 January, 2014).
- Bohn, K., Knaus, J., Wiese, R., and Domahs, U. (2013). The influence of rhythmic (ir)regularities on speech processing: evidence from an ERP study on German phrases. *Neuropsychologia* 51, 760–771. doi: 10.1016/j.neuropsychologia.2013.01.006
- Brennan, C., Cao, F., Pedroarena-Leal, N., McNorgan, C., and Booth, J. R. (2013). Reading acquisition reorganizes the phonological awareness network only in alphabetic writing systems. *Hum. Brain Mapp.* 34, 3354–3368. doi: 10.1002/hbm.22147
- Bulheller, S., and Häcker, H. O. (2002). *Coloured Progressive Matrices (CPM). Deutsche Bearbeitung und Normierung nach J. C. Raven*. Frankfurt: Pearson Assessment.
- Castro-Caldas, A., Petersson, K. M., Reis, A., Stone-Elander, S., and Ingvar, M. (1998). The illiterate brain. Learning to read and write during childhood influences the functional organization of the adult brain. *Brain* 121, 1053–1063. doi: 10.1093/brain/121.6.1053
- Cooper, N., Cutler, A., and Wales, R. (2002). Constraints of lexical stress on lexical access in English: evidence from native and non-native listeners. *Lang. Speech* 45, 207–228. doi: 10.1177/00238309020450030101
- Desroches, A. S., Cone, N. E., Bolger, D. J., Bitan, T., Burman, D. D., and Booth, J. R. (2010). Children with reading difficulties show differences in brain regions associated with orthographic processing during spoken language processing. *Brain Res.* 1356, 73–84. doi: 10.1016/j.brainres.2010.07.097
- Escudero, P., and Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Lang. Speech* 53, 343–365. doi: 10.1177/0023830910371447
- Friedrich, C. K. (2005). Neurophysiological correlates of mismatch in lexical access. *BMC Neurosci.* 6:64. doi: 10.1186/1471-2202-6-64
- Friedrich, C. K., Kotz, S. A., Friederici, A. D., and Alter, K. (2004). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Cogn. Brain Res.* 20, 300–308. doi: 10.1016/j.cogbrainres.2004.03.007
- Friedrich, C. K., Lahiri, A., and Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: asymmetries with word initial variations. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1545–1559. doi: 10.1037/a0012481
- Friedrich, C. K., Schild, U., and Röder, B. (2009). Electrophysiological indices of word fragment priming allow characterizing neural stages of speech recognition. *Biol. Psychol.* 80, 105–113. doi: 10.1016/j.biopsycho.2008.04.012
- Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., and Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56, 1127–1134. doi: 10.1016/j.neuron.2007.09.038
- Giraud, A. L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063
- Goswami, U., and Bryant, P. E. (1990). *Phonological Skills and Learning to Read*. Hillsdale, NJ: Erlbaum.
- Goswami, U., Ziegler, J. C., and Richardson, U. (2005). The effects of spelling consistency on phonological awareness: a comparison of English and German. *J. Exp. Child Psychol.* 92, 345–365. doi: 10.1016/j.jecp.2005.06.002
- Grainger, J., and Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: cross-task comparisons. *J. Mem. Lang.* 35, 623–647. doi: 10.1006/jmla.1996.0033
- Grainger, J., and Holcomb, P. J. (2009). Watching the word go by: on the time-course of component processes in visual word recognition. *Lang. Linguist. Comp.* 3, 128–156. doi: 10.1111/j.1749-818X.2008.00121.x
- Jansen, H., Mannhaupt, G., Marx, H., and Skowronek, H. (2002). *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC) (2., überarb. Aufl.)*. Göttingen: Hogrefe
- Lenhard, W., and Schneider, W. (2006). *ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe.
- Lieberman, M., and Prince, A. (1977). On stress and linguistic rhythm. *Linguist. Inq.* 8, 249–336.

- Luo, H., and Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Front. Psychol.* 3:170. doi: 10.3389/fpsyg.2012.00170
- Lyytinen, H., Aro, M., Eklund, K., Erskine, J., Guttorm, T., Laakso, M. M., et al. (2004). The development of children at familial risk for dyslexia: birth to early school age. *Ann. Dyslexia* 54, 184–220. doi: 10.1007/s11881-004-0010-3
- Mitterer, H., and McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE* 4:e7785. doi: 10.1371/journal.pone.0007785
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time.” *Speech Commun.* 41, 245–255. doi: 10.1016/S0167-6393(02)00107-3
- Poeppel, D., Idsardi, W. J., and van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1071–1086. doi: 10.1098/rstb.2007.2160
- Reinisch, E., Jesse, A., and McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: lexical stress drives eye movements immediately. *Q. J. Exp. Psychol. (Colchester)* 63, 772–783. doi: 10.1080/17470210903104412
- Schild, U., Röder, B., and Friedrich, C. K. (2011). Learning to read shapes the activation of neural lexical representations in the speech recognition pathway. *Dev. Cogn. Neurosci.* 1, 163–174. doi: 10.1016/j.dcn.2010.11.002
- Schild, U., Röder, B., and Friedrich, C. K. (2012). Neuronal spoken word recognition: the time course of processing variation in the speech signal. *Lang. Cogn. Process.* 27, 159–183. doi: 10.1080/01690965.2010.503532
- Soto-Faraco, S., Sebastián-Gallés, N., and Cutler, A. (2001). Segmental and Suprasegmental Mismatch in Lexical Access. *J. Mem. Lang.* 45, 412–432. doi: 10.1006/jmla.2000.2783
- Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., et al. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *J. Neurosci.* 29, 14726–14733. doi: 10.1523/JNEUROSCI.1246-09.2009
- Tomlinson, J. M., Liu, Q., and Fox Tree, J. E. (2013). The perceptual nature of stress shifts. *Lang. Cogn. Process.* 1–13. doi: 10.1080/01690965.2013.813561
- van Donselaar, W., Koster, M., and Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Q. J. Exp. Psychol. A* 58, 251–273. doi: 10.1080/02724980343000927
- Ventura, P., Kolinsky, R., Pattamadilok, C., and Morais, J. (2008). The developmental turnpoint of orthographic consistency effects in speech recognition. *J. Exp. Child Psychol.* 100, 135–145. doi: 10.1016/j.jecp.2008.01.003
- Ventura, P., Morais, J., and Kolinsky, R. (2007). The development of the orthographic consistency effect in speech recognition: from sublexical to lexical involvement. *Cogn. Int. J. Cogn. Sci.* 105, 547–576. doi: 10.1016/j.cognition.2006.12.005
- Walley, A., Metsala, J., and Garlock, V. M. (2003). Spoken vocabulary growth: its role in the development of phoneme awareness and early reading ability. *Read. Writ.* 16, 5–20. doi: 10.1023/A:1021789804977
- Witteman, J., van Ijzendoorn, M. H., van de Velde, D., van Heuven, V. J. J. P., and Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: a meta-analysis of the lesion literature. *Neuropsychologia* 49, 3722–3738. doi: 10.1016/j.neuropsychologia.2011.09.028
- Zatorre, R. J., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953. doi: 10.1093/cercor/11.10.946
- Zecker, S. G. (1991). The orthographic code: developmental trends in reading-disabled and normally-achieving children. *Ann. Dyslexia* 41, 178–192. doi: 10.1007/BF02648085
- Ziegler, J. C., and Ferrand, L. (1998). Orthography shapes the perception of speech: the consistency effect in auditory word recognition. *Psychon. Bull. Rev.* 5, 683–689. doi: 10.3758/BF03208845
- Ziegler, J. C., and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychol. Bull.* 131, 3–29. doi: 10.1037/0033-2909.131.1.3

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 November 2013; accepted: 13 May 2014; published online: 03 June 2014.  
Citation: Schild U, Becker ABC and Friedrich CK (2014) Processing of syllable stress is functionally different from phoneme processing and does not profit from literacy acquisition. *Front. Psychol.* 5:530. doi: 10.3389/fpsyg.2014.00530  
This article was submitted to Language Sciences, a section of the journal *Frontiers in Psychology*.  
Copyright © 2014 Schild, Becker and Friedrich. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# The direction of word stress processing in German: evidence from a working memory paradigm

Frank Domahs<sup>1\*</sup>, Marion Grande<sup>2</sup>, Walter Huber<sup>2</sup> and Ulrike Domahs<sup>1,3</sup>

<sup>1</sup> Klinische Linguistik, Institut für Germanistische Sprachwissenschaft, Philipps-Universität Marburg, Marburg, Germany

<sup>2</sup> RWTH Aachen University Hospital, Aachen, Germany

<sup>3</sup> University of Cologne, Cologne, Germany

## Edited by:

Richard Wiese, Philipps-Universität Marburg, Germany

## Reviewed by:

Claudia K. Friedrich, University of Tübingen, Germany

Birgit Alber, Università degli Studi di Verona, Italy

## \*Correspondence:

Frank Domahs, Klinische Linguistik, Institut für Germanistische Sprachwissenschaft, Philipps-Universität Marburg, Wilhelm-Röpke-Straße 6A, D-35032 Marburg, Germany  
e-mail: domahs@uni-marburg.de

There are contradicting assumptions and findings on the direction of word stress processing in German. To resolve this question, we asked participants to read tri-syllabic non-words and stress ambiguous words aloud. Additionally, they also performed a working memory (WM) task (2-back task). In non-word reading, participants' individual WM capacity was positively correlated with assignment of main stress to the antepenultimate syllable, which is most distant to the word's right edge, while a (complementary) negative correlation was observed with assignment of stress to the ultimate syllable. There was no significant correlation between WM capacity and stress assignment to the penultimate syllable, which has been claimed to be the default stress pattern in German. In reading stress ambiguous words, a similar but non-significant pattern was observed as in non-word reading. In sum, our results provide first psycholinguistic evidence supporting leftward stress processing in German. Our results do not lend support to the assumption of penultimate default stress in German. A specification of the lemma model is proposed which seems able to reconcile our findings and apparently contradicting assumptions and evidence.

**Keywords:** lexical stress, directionality, reading, non-words, pseudowords, correlation

## INTRODUCTION

How do we know which syllable of a polysyllabic word should receive main stress? In figuring out – should we start from the beginning (i.e., left edge) or from the end (i.e., right edge) of the word? The answer seems easy in languages with fixed stress position: We should start from the left edge in languages with fixed stress on the first (e.g., Cahuilla, Hungarian, and Icelandic), second (e.g., Dakota, Mapudungun, and Tolai), or third (e.g., Winnebago) syllable, while we should start from the right edge in languages with fixed stress on the ultimate (U, e.g., Balinese, Persian, and Weri), penultimate (PU, e.g., Djingili, Polish, and Quechua), or antepenultimate (APU, e.g., Greek, Macedonian, and Paumari) syllable (for an overview see Goedemans and van der Hulst, 2014). The matter is less obvious in languages with variable stress (e.g., English, German, and Russian). As in those languages the position of main stress is largely unpredictable, it has been suggested that this information has to be stored in the mental lexicon for all words. However, it is not clear, whether, for instance, the lexical entry of the German word *Veranda* codes main stress position as second or prefinal – in other words, whether retrieval of stress position proceeds in a rightward or leftward manner (or with no specific directionality at all).

Based on regularities or analogies generated from their lexical knowledge, even speakers of languages with unpredictable stress are able to assign stress to non-words (Janssen, 2003b; Tappeiner et al., 2007; Röttger et al., 2012; Domahs et al., 2014). Typically, the assignment of stress to non-words is characterized by large interindividual variance. Moreover, the assignment of stress to both existing words and non-words may leave behavioral traces

of processing demands. The present study aims to explore the interaction of interindividual variance in stress assignment and specific computational demands for different stress positions to investigate the direction of stress processing in German. In the remainder of this section, we will first summarize arguments on the direction of stress computation in German and then outline the rationale of the study.

## THE DIRECTION OF STRESS COMPUTATION GERMAN

The computation of main stress position may, in principle, start from the beginning or from the end of the word (i.e., rightward or leftward assignment, respectively). There are arguments for both options in German, which will be reviewed in the following.

Most current accounts on German stress assignment – explicitly or implicitly – proceed from the assumption that the syllable to be assigned main stress is defined in a leftward fashion, starting from the right edge of a word. This holds true irrespective of whether these accounts opt for quantity-sensitive or for quantity-insensitive stress assignment. Quantity-sensitive accounts state that the structure or weight of the final and/or prefinal syllable is a particularly important predictor of the position of main stress in German (Vennemann, 1991; Féry, 1998; Domahs et al., 2008). Quantity-insensitive accounts typically assume that the PU is the default stress position in German, all other stress patterns being exceptions which require lexicalization (Eisenberg, 1991; Wiese, 1996). Leftward stress computation in German is supported by the fact that only one of the last three syllables (APU, PU, or U) can bear main stress (“three-syllable window,” Giegerich, 1985; Vennemann, 1991; Zonneveld

et al., 1999). Psychologically, the three-syllable window seems to be very robust. It was, for example, obeyed in a patient with acquired language impairment, who otherwise showed severe phonological and prosodic deficits (Janßen and Domahs, 2008).

One major argument for rightward stress computation comes from the psycholinguistic “lemma” model of speech production developed by Levelt et al. (1999). In this model, a metrical frame is retrieved (independently from the sequence of phonemes) which determines the number of syllables and the position of main stress in case of non-default stress assignment. In a further processing step, which is called prosodification, the metrical frame is filled with segments and – in the case of default assignment – the stress position is assigned. Crucially, first syllable stress is assumed to be the default in German (as in Dutch and English). In fact, evidence reported by Schiller et al. (2006) seems to lend support to such a rightward processing of metrical stress in Dutch: In a monitoring task, subjects were faster to detect stressed syllables at the beginning compared to stressed syllables at the end of words which they had to name implicitly from pictures. Yet, these authors themselves note that their observation may also be caused by the incremental (i.e., rightward) functioning of the monitoring system rather than by the incremental functioning of stress processing itself. Note that the assumption of first syllable stress as default in German as implemented in the lemma model (Levelt et al., 1999), although conceptually in clear contrast to the assumption of a PU default in German (Eisenberg, 1991; Wiese, 1996), makes identical predictions for the huge bulk of existing words, given that most monomorphemic word types in the German corpus consist of one or two syllables. In trisyllabic words, however, the predictions based on leftward computation differ from the predictions of the lemma model.

There is a third set of accounts, which assume that there are two co-phonologies of German with different implications for the direction of stress computation. According to those accounts, the default position of main stress in native German words is the first syllable, whereas stress in non-native words would be computed in a leftward manner starting from the right edge of the word (Wurzel, 1970, 1980; Benware, 1980; Féry, 1986). However, a number of authors disagree with the need to distinguish between native and non-native German phonology (Giegerich, 1985; Hall, 1992; Wiese, 1996).

## THE PRESENT STUDY

In sum, the question in which direction metrical stress is computed in German is still open. In our experiment, we explored the possibility that the processing of word stress in German occurs in a leftward instead of a rightward fashion, as predicted by a number of different phonological theories (Eisenberg, 1991; Vennemann, 1991; Wiese, 1996; Féry, 1998; Domahs et al., 2008). Note that we are taking a cognitive perspective here, rather than a purely descriptive linguistic approach. In this cognitive perspective, different stress positions may be associated with different computational costs. Specifically, processing costs, operationalized as working memory (WM) load, should increase with increasing distance of stress position from the starting point of computation

(left or right edge of the word). If stress computation works from right to left, then computational WM load should increase in the following direction: U < PU < APU stress position. The opposite hierarchy is expected in the case of rightward stress computation. For instance, the assignment of stress to the first and the final syllable in non-words with a VC.V.VCC<sup>1</sup> structure (e.g., *Rulkomenk*) is approximately balanced across participants in group analyses (43 and 47%, respectively), while the second syllable is only rarely stressed (Janssen, 2003b). However, if the computation of stress operates, indeed, in a leftward fashion, then it requires additional processing steps to identify and stress the APU position compared to placing stress on the U position, i.e., APU stress assignment is computationally more demanding than U stress assignment. Consistent with the right-to-left hypothesis, two patients with reduced WM span (Janssen, 2003a; Janßen and Domahs, 2008) produced virtually no APU stress on pseudowords, while a group of healthy subjects produced up to 50% of this stress pattern with the same material. More generally, the existence of the three-syllable window in German may be interpreted as consequence of leftward stress assignment subject to processing limitations.

To pursue our hypothesis, we examined non-word reading of native speakers of German whose WM capacity was quantified using a 2-back task (Zimmermann and Fimm, 1993). The use of non-words not only avoids the influence of lexical variables (e.g., word frequency) as far as possible but also ensures that the stress position has to be computed instead of retrieved from long-term memory (i.e., the mental lexicon).

We wanted to use the fact that there is a large degree of interindividual heterogeneity in word stress assignment, at least partly related to WM (Heisterueber et al., 2014). Specifically, it was predicted that the proportion of computationally complex APU stress assignment across stimuli should be positively correlated with the individual WM capacity. In other words: the more limited the WM capacity the fewer computationally complex stress assignments should be observed.

Participants were also asked to read a short story containing words, which can be stressed on different syllables (i.e., stress ambiguous words). Given that German is a language with largely unpredictable stress, the position of main stress should be lexicalized for these words. However, it may still be that a participant's WM capacity influences his/her preferred stress position for such words. This influence may be less strong than the one expected for non-words, as the computational impact of stress position may be less pronounced in lexical retrieval than in actual computation of a stress pattern.

Some accounts of stress assignment in German assume that tri-syllabic words with a closed final syllable are parsed into two metrical feet (a final non-branching foot and a preceding binary one ( $[\sigma(\sigma)_F(\sigma)_F]_\omega$ ), while words with an open final syllable are only parsed into one foot ( $[\sigma(\sigma)_F]_\omega$ ), leaving an unparsed initial syllable, where stress assignment is disfavored (Alber, 1997; Domahs et al., 2008; Knaus and Domahs, 2009). Based on this

<sup>1</sup>Here and in the remainder of this paper syllable structure only indicates vowels (V) and consonants (C) of the syllabic rhyme. Syllabic onset structure may vary and is therefore not specified.



analysis, words with a closed final syllable have more potential landing sites for main stress than words with open final syllable.

Note that the potential existence of a default stress pattern may overwrite the effect of computational direction. In this case, it may be that the default stress assignment is computationally easier than stress assignment to other positions. Potential default stress positions in German are the first syllable (Levelt et al., 1999) or the PU (Eisenberg, 1991; Wiese, 1996).

In sum, we want to make use of interindividual variance in cognitive processing capacity to distinguish between easier and more difficult stress positions indicative of the direction of stress computation in German.

## MATERIALS AND METHODS

### PARTICIPANTS

Participants were recruited from retirement homes in the city of Aachen (Germany) and the orthopedic ward of the RWTH Aachen University Hospital. Thirty-eight participants performed a reading task with a list of 60 existing German words (20 words with APU, PU, and U stress pattern, respectively, in randomized order, see Tappeiner et al., 2007). Two participants read less than 80% correct and were excluded from further analyses. In the remaining sample, there were 20 women and 16 men. All participants were native speakers of German, coming from a heterogeneous educational background (6 had obtained German *Abitur*, 19 had finished *Realschule*, and the remaining 11 had finished *Hauptschule*). All but two participants were right handed according to their own disclosure.

Participants were aged between 52 and 94 years (mean = 72.1). This age range was chosen to increase the interindividual variance in WM capacity (Dobbs and Rule, 1989; Brockmole and Logie, 2013; Murre et al., 2013). No participants with diagnosed dementia or neurological illness were included. It was made sure that all participants used their glasses and/or hearing aid if necessary. All participants gave their informed consent and received a compensation of 5 Euros. The study was approved by the Institutional Review Board of the Medical Faculty at RWTH Aachen University (EK 182/06).

### TASKS

Participants performed three tasks: a non-word reading task, a reading task with stress-ambiguous existing words, and a 2-back task.

We used the non-word reading task designed by Janssen (2003b; see also Domahs et al., 2014). In this task, non-words have to be produced within a carrier sentence, to prevent from artifacts due to reading isolated non-words in a list. The carrier sentence was always the same throughout the task (*Ich habe gehört, dass Peter ... gesagt hat. [I have heard that Peter said ...]*) to control for interference from sentence prosody. Participants were instructed to first read the non-word silently and only if they felt ready to produce it fluently to utter the carrier sentence containing the target-non-word.

In the second task, participants were asked to read a small purpose-made story (33 lines) containing 8 existing words which are stress ambiguous in German (see Stimuli). Target words were

not highlighted in the text and participants were unaware of the specific purpose of this task (i.e., they were globally instructed to read the story aloud).

The 2-back task is a supplement to a larger battery testing attentional functions (TAP, Zimmermann and Fimm, 1993). Participants see a sequence of isolated letters on a screen, in a self-paced speed of presentation. They are asked to indicate by a button press, whether any given letter in this sequence is identical to its pre-predecessor. (e.g., A–E–C–E–K–L–K, required yes-answers highlighted). Thus, this demanding task requires a variety of executive or WM functions including storing and updating of relevant information and inhibition of irrelevant information. As a kind of shorthand term, we will refer to the underlying construct tested with the 2-back task as WM capacity. Note that in this task the position of elements within a sequence is crucially important.

### STIMULI

In the non-word reading task, we used the set of stimuli described by Janssen (2003b, see also Domahs et al., 2014). These are phonotactically legal three-syllabic non-words in eight syllable structure conditions (rhyme structures: VC.V.VCC, V.VC.VCC, VC.V.VC, V.V.VC, V.VC.VC, V.V.V, V.VC.V, and VC.VC.V). These eight conditions were designed to examine the role of syllable structure on stress assignment, particularly focusing on the weight of the final syllable, as this seems to be most influential (Janssen, 2003b; Röttger et al., 2012; Domahs et al., 2014). However, the stimulus set did not include all logically possible combinations of syllable structures. Conditions with three heavy syllables (VC.VC.VC and VC.VC.VCC) were excluded, because such words are not attested in German. Furthermore, words with super-heavy syllables and light penult and antepenultimate (V.V.VCC) as well as with light final and penult and heavy antepenult (VC.V.V) were not tested, because such conditions would probably not add further insights into the role of quantity on stress assignment. In the item construction, resyllabifications of coda consonants as onset consonants of the following syllable were avoided by filling each onset position. In addition, in syllable contacts the sonority of consonants was chosen such that the parsing of consonants into complex onsets was made unlikely (e.g., a non-word like *bat.ram* could be syllabified as *ba.tram*, while *las.fon.ta* cannot be syllabified as *\*la.sfon.ta*). Potential similarities to existing words were avoided as far as possible by including only items whose final two syllables did not rhyme with existing words. For further details on stimulus selection, see Domahs et al. (2014).

There were 10 items per condition, 80 items overall, which were presented in pseudorandomized order, interspersed with 40 one- and two-syllable filler non-words as well as 13 four-syllable non-words to prevent participants from using an individual “default” stress pattern consistently across the whole list of items. Although in general the target non-words lead to different specific stress assignment preferences depending on their syllable structure (e.g., words with V.VC.V structure are preferably stressed on PU syllable), there is always a large degree of interindividual variance – which so far is left unexplained – such that in no condition non-words are exclusively stressed on one syllable (Janssen, 2003b; Tappeiner et al., 2007; Röttger et al., 2012).



In the word reading task, eight target words embedded in a short story were stress ambiguous, i.e., they can receive either APU or U stress in German (*Kabarett, Telefon, Mikroskop, Dromedar, Marzipan, Alkohol, Megafon, Horizont*). Stress ambiguity was confirmed by the Duden® online dictionary ([www.duden.de](http://www.duden.de)). The fact that most stress ambiguities in German involve APU vs. U main stress position can be accounted for by the similarity of their underlying foot structure and the dissimilarity of the underlying PU foot structure and by the related fact that only words with APU and U stress consist of two metrical feet and therefore allow for stress variance (Domahs et al., 2008, 2013; Janßen and Domahs, 2008; Röttger et al., 2012). Note that preference for a specific variant of stress ambiguous words largely depends on the speaker's regional variant of German. Other possible sources of interindividual variance, in particular WM capacity, have not been reported so far.

## ANALYSES

Participants' oral responses were recorded and transcribed later by a trained speech-language therapist who was blind to the hypotheses of the experiment. Main stress was determined based on perceptual judgment<sup>2</sup>. In cases of any uncertainty, transcription was discussed with an experimenter. If no consensus could be obtained for a specific item, this item was excluded from analyses.

In non-word reading, only responses without segmental errors for which main stress position could be identified unambiguously were included in the analyses. These criteria were fulfilled by 90.8% of the given responses. Dependent variables were the proportions of APU, PU, and U stress assignment in the target non-words. Note that these proportions are interdependent.

In word reading, the dependent variable was the proportion of APU stress assignment in the target words. Note that the proportion of U stress assignment is complementary with the proportion of APU stress assignment. There were 94.1% analyzable word items.

In the 2-back task, the dependent variable ("WM capacity") was the number of correct yes-responses ( $\max = 14$ ).

Given the non-parametric nature of the data, we explored the relationship between individual WM capacity on the one hand and the proportion of stress patterns in non-word and word reading on the other using Spearman's rank correlation coefficient.

## RESULTS

### NON-WORDS

We found a significant positive correlation between individual WM capacity and the proportion of APU stress assigned ( $r_s = 0.344$ ,  $p = 0.040$ ) and a (complementary) negative correlation between WM capacity and the proportion of final stress assigned ( $r_s = -0.427$ ,  $p = 0.009$ ; see **Table 1**, **Figure 1**). There was no significant correlation between WM

capacity and the proportion of PU stress assigned ( $r_s = -0.081$ ,  $p = 0.637$ ).

The pattern of correlations was consistent across conditions (i.e., positive for APU, negative for U, and non-significant for PU). However, looking at the influence of syllable structure (possibly indicative of foot structure), significant correlations were almost exclusively found for non-words with closed final syllable, which – due to their foot structure – offer more potential landing sites for main stress than non-words with open final syllable (see **Table 1**, conditions 1–5).

Moreover, there was increasing interindividual variance of APU stress assignment with increasing WM capacity, i.e., there were increasing absolute residuals from a linear function ( $r_s = 0.352$ ,  $p = 0.035$ ). There were no such significant relationships between WM capacity and the variance of either PU ( $r_s = 0.045$ ,  $p = 0.796$ ) or U ( $r_s = 0.260$ ,  $p = 0.126$ ) stress assignment.

### STRESS AMBIGUOUS WORDS

There was a near-significant negative correlation between WM capacity and the proportion of final stress assigned ( $r_s = -0.321$ ,  $p = 0.057$ ), but no significant correlation between WM capacity and the proportion of APU stress assigned ( $r_s = 0.195$ ,  $p = 0.254$ ; see **Figure 2**). Note that these correlations are not completely complementary due to 5.9% unanalyzable trials.

## DISCUSSION

In sum, we observed a positive correlation of WM capacity with the proportion of APU stress assigned and a (complementary) negative correlation with the proportion of U stress assigned for non-words and a similar but non-significant pattern also for stress ambiguous words. There was no correlation of WM capacity with PU stress assignment. We would like to argue that this pattern of results speaks in favor of a leftward processing of word stress in German as participants with limited WM capacity only rarely produced the computationally most demanding (i.e., most leftward) APU pattern, while participants with good WM capacity were able to use APU stress. This interpretation is also supported by the increasing variance of APU stress assignment with increasing WM capacity: while participants with good WM were well able to assign APU stress, they were not restricted to that pattern.

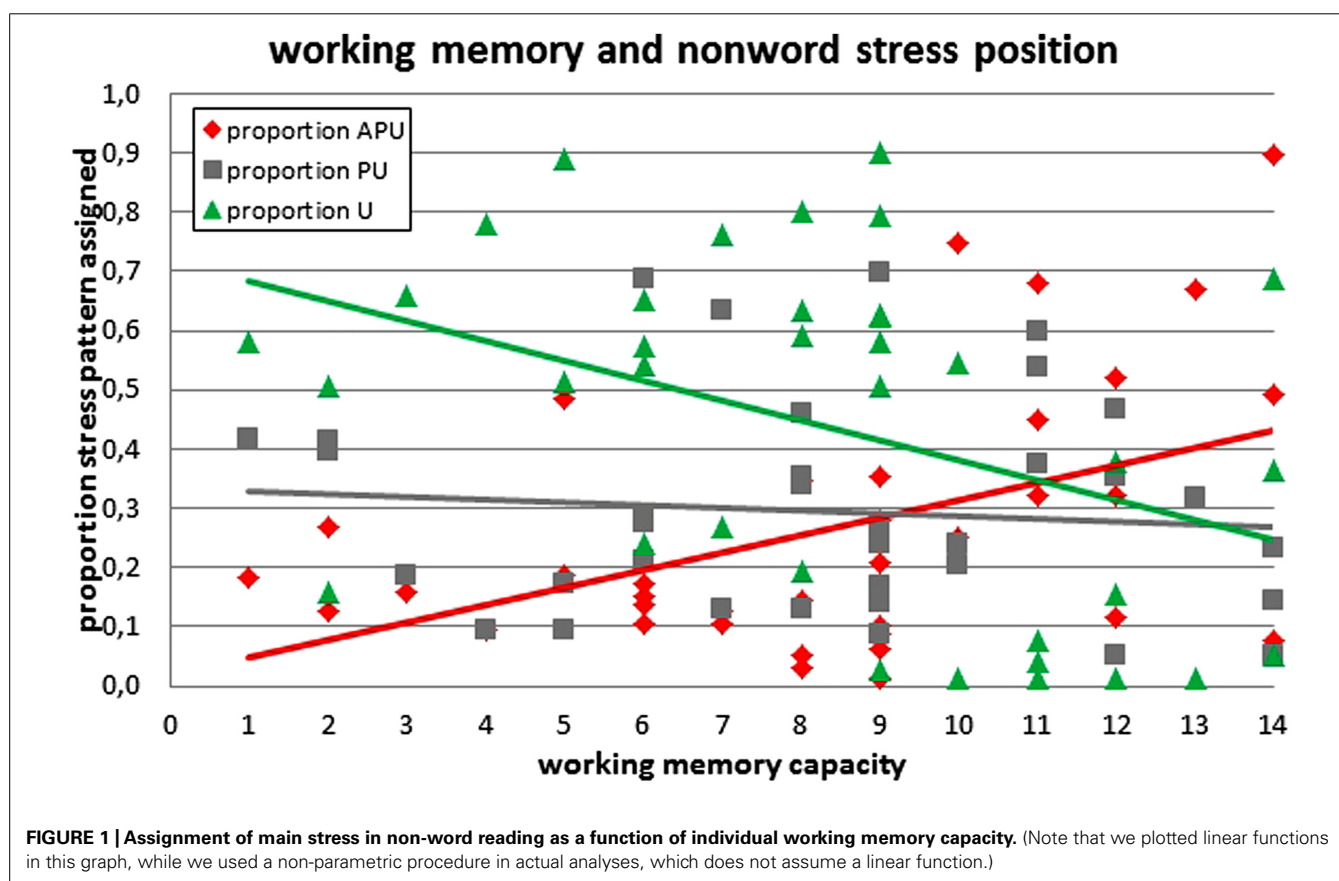
More specifically, our observation that correlations were almost exclusively found for non-words with a closed final syllable is consistent with the assumption that tri-syllabic words with a closed final syllable are parsed into two metrical feet (a final non-branching foot and a preceding binary one) while words with an open final syllable are only parsed into one foot, leaving an unparsed initial syllable (Alber, 1997; Domahs et al., 2008; Knaus and Domahs, 2009). The two metrical feet, that words with closed final syllable are made of, offer two potential positions for main stress, while tri-syllabic words with open final syllable typically consist of only one metrical foot and an unparsed initial syllable, where main stress should be disfavored. In consequence, words with one metrical foot provide only one option of stress assignment, whereas words with two feet do require a decision where to place main stress, which may lead to increased processing costs. Indeed, non-words with open final syllable tended to attract less APU stress than words with closed final syllable, consistent with

<sup>2</sup>Although perceptual judgment is the standard method in stress assignment experiments using production paradigms, this method does not allow disentangling the potentially distinct contribution of the different phonetic cues (e.g., pitch, duration, and intensity) which – in a complex interplay – lead to the perception of stress.

**Table 1 | Results as a function of structural conditions.**

	Condition	Correlations			Proportion (%)			
		APU	PU	U	APU	PU	U	Not analyzable
1	VC.V.VCC	0.393*	-0.133	-0.316	30.8	14.2	42.8	12.2
2	V.VC.VCC	0.535***	0.084	-0.404*	21.2	24.4	42.8	11.7
3	VC.V.VC	0.357*	0.078	-0.369*	33.9	13.3	43.9	8.9
4	V.V.VC	0.385*	-0.176	-0.353*	36.1	13.9	42.8	7.2
5	V.VC.VC	0.282	-0.082	-0.243	20.3	24.4	44.2	11.1
6	V.V.V	0.319	-0.125	-0.190	17.5	43.1	33.6	5.8
7	V.VC.V	0.046	0.184	-0.304	15.0	42.8	32.5	9.7
8	VC.VC.V	0.009	0.104	-0.361*	17.5	43.9	31.4	7.2
	Total	0.344*	-0.081	-0.427**	24.0	27.5	39.2	9.2

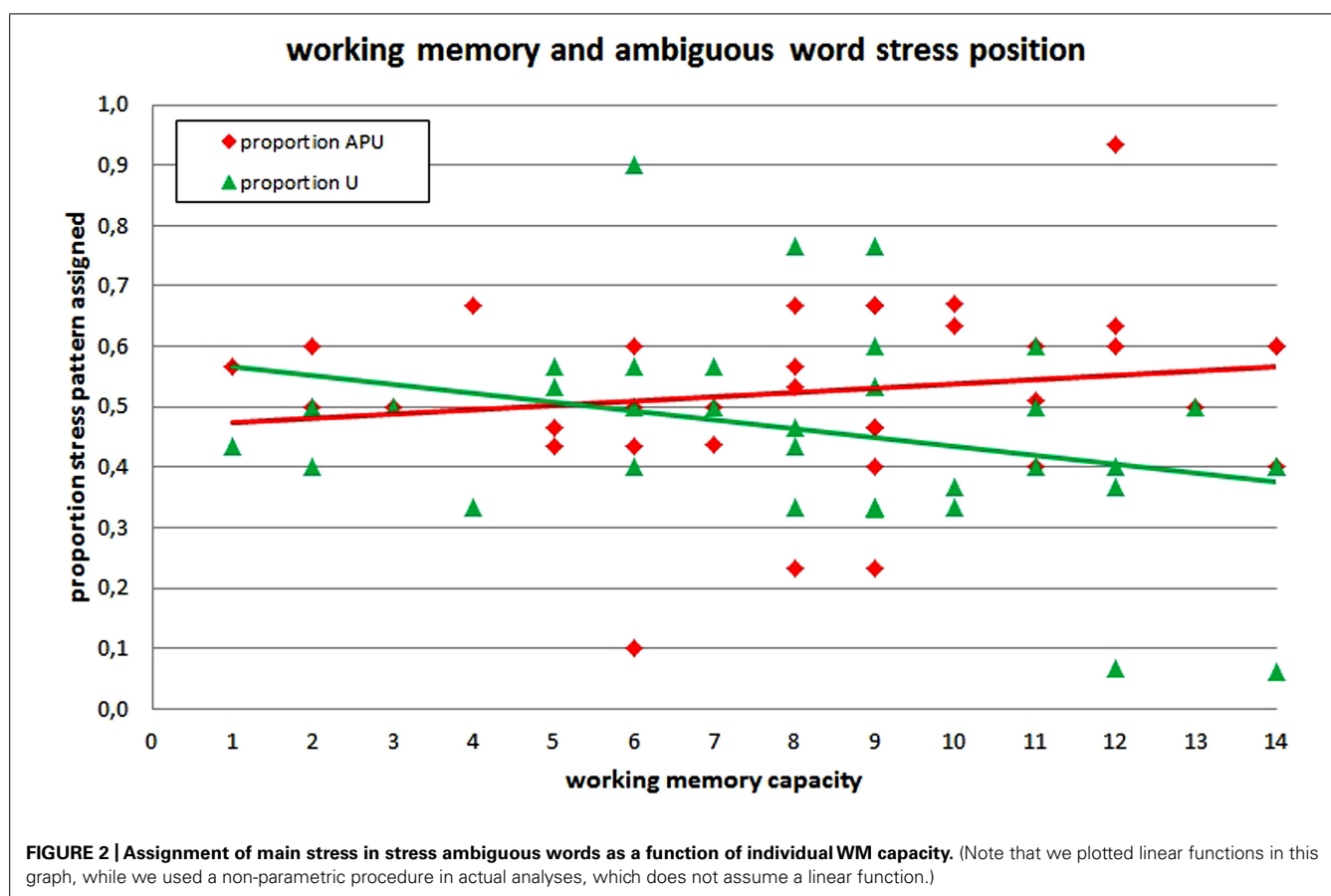
Correlations are indicated as Spearman's rank correlation coefficients between WM capacity and proportion of stress position assigned. Significant correlations are marked with \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , or \*\*\* $p \leq 0.001$ , respectively.



previous findings (Janssen, 2003b; Röttger et al., 2012; Domahs et al., 2014).

It may be argued that leftward stress computation leads to increased costs in speech production compared to rightward stress computation. Given that the sequence of phonemes is processed in a rightward manner, rightward stress processing is consistent with the processing direction of phonemes whereas leftward stress

processing is inconsistent with it, causing elevated costs. Similar arguments have been put forward for left aligning vs. right aligning systems of secondary stress (Hayes, 1995; Alber, 2005). In left aligning systems, less phonological pre-planning may be required in speaking, given that the parser does not have to know the number of a word's syllables before starting to assign stress. With respect to main stress, a look at the World Atlas of Language



Structures Online (Goedemans and van der Hulst, 2014; features 14A and 15A) reveals that systems with right edge orientation are not rare. If, indeed, such systems are associated with increased processing costs, it remains an open question whether there is any compensation for this disadvantage in languages with right edge orientation including German.

If PU stress should be regarded as default option in German (Eisenberg, 1991; Wiese, 1996), one may have expected a processing advantage such that participants with limited WM capacity assign more (“default”) PU stress than participants with good WM capacity. Obviously, this was not the case. Previous studies have also failed to provide empirical evidence for a processing advantage of PU stress in German. PU stress was not the preferred pattern in violation paradigms (Domahs et al., 2008, 2013). Moreover, it was not dominant in monolingual (Röttger et al., 2012) or bilingual (Tappeiner et al., 2007) non-word production experiments either. Finally, PU stress was not the most robust pattern in cases of acquired language impairment (Janssen, 2003a; Janßen and Domahs, 2008). Clearly, we did not find evidence for a first syllable default either.

In sum, we would like to explain the present pattern of results based on cognitive procedures which operate in a leftward fashion to assign stress to non-words. These procedures may be sets of rules or constraints (Domahs et al., 2008; Knaus and Domahs, 2009), while we found no evidence for the psychological reality of a default stress position in German. For the first time, it

has been demonstrated that the use of these procedures is influenced by individual cognitive processing capacity. At present, the procedures which are actually used during the computation of main stress remain unspecified. Nonetheless we suggest so far that (a) it seems to be more demanding to assign stress to a syllable (APU) which is distant from the starting point of these procedures (i.e., from the right edge of a word) than to a syllable (U) which is close to it and (b) this difference is more pronounced in non-words which contain two metrical feet compared to non-words containing only one foot. Future work should try to further elucidate the exact nature of stress assigning algorithms.

A processing-based account of stress assignment as sketched above could also explain the observation that two patients with impaired lexical knowledge due to primary progressive aphasia did not use any APU stress in cases of uncertainty (Janssen, 2003a; Janßen and Domahs, 2008). Given that both patients had a massively reduced WM span, APU stress assignment may have been too demanding for them. Their avoidance of APU stress has previously been explained with APU stress being exceptional and needing to be lexicalized (Knaus and Domahs, 2009). However, in the light of the present findings the processing-based account seems to be superior: It was the good participants who produced the largest proportion of the putative exceptional pattern and the participants with limited WM capacity who tended to avoid it.

We also found a near-significant negative correlation between WM capacity and the proportion of stress assignment to the final syllable for stress ambiguous words. This is remarkable given that stress assignment to German words is assumed to be fully lexicalized. In the case of stress ambiguous words, the two variants of a word (e.g., *Hórizont* vs. *Horizónt*) show a regional distribution, while individual speakers stick to one of the variants quite consistently, i.e., they have one lexical entry which is determined by their regional variant of German. However, the present results suggest that this is not the whole story. There was interindividual variance in stress assignment to those words which was systematically related to individual WM capacity while all participants were recruited in the same area (Aachen, Germany). In our view, there are two possible explanations for processing-related interindividual variance in stress assignment to ambiguous words: first, it may be that lexical retrieval of word stress in German has some form of right alignment. Thus, for some reason or another, the scanning of lexical entries for their stress information may occur in a leftward manner. Second, it may be that lexical retrieval co-occurs with the application of (rule-based) procedures as applied for stress assignment in non-words and that both processes interact. Given that at least one of the two processes (i.e., rule application) is more demanding for APU than for U stress, this may also influence lexical retrieval or the final articulatory output – at least in stress ambiguous words. Note, however, that all participants were well able to read existing unambiguous words correctly such that they produced merely any variance in stress assignment in such words.

Could the increasing use of ultimate stress with decreasing WM capacity be explained by factors other than the computational ease to place main stress on the final syllable? An alternative explanation may refer to articulatory preparation in general rather than to the computation of main stress position. According to this explanation, participants with limited WM span might tend to lengthen the final syllable of target (non)words to get more time for preparing the subsequent word and – given that duration is also a relevant phonetic cue to word stress – this might result in perceived stress on the final syllable. However, this potential artifact has been minimized in our experimental design as our target stimuli were embedded in carrier sentences/text, respectively. Recall that the carrier sentence was identical for all non-words, such that the word following the target was highly expectable and automatized during the experiment. Moreover, in non-word reading participants were instructed to first read the non-word silently and only if they felt ready to produce it fluently to utter the carrier sentence containing the target-non-word. Yet, to clear away the last doubts and to disentangle ultimate stress from final lengthening, further research may address languages with rightward stress assignment, where both effects would be separable.

Our data support accounts of German word stress, which assume leftward computation (e.g., Vennemann, 1991; Alber, 1997; Féry, 1998; Domahs et al., 2008), but seem to be at odds with those assuming rightward computation (Levelt et al., 1999). However, our data are silent with respect to the possibility that there are two co-phonologies of German with different stress computation directions (Wurzel, 1970, 1980; Benware, 1980; Féry, 1986). Given

that we used tri-syllabic non-words in one task and that the stress ambiguous words used in the other task were all loan words, it seems plausible that both types of stimuli were treated within the “non-native” phonology by our participants. In this case, results of both tasks would lend support to the assumption that in this part of German phonology, stress computation proceeds in a leftward manner. Note that experimental evidence based on processing difficulty would be difficult to obtain for the other part of German phonology (native words), as this comprises mainly one- and two-syllable words which do not offer the possibility of “long distance” stress assignment.

In the remainder of this section we would like to argue that it is possible to reconcile the apparently contradicting assumptions on the direction of stress processing in German, based on a specification of the lemma model of speech production (Levelt et al., 1999). Recall that within this model, there are two stages of prosodic encoding: at the first stage (*frame generation*), a metrical frame is generated, specifying the number of syllables and the stress pattern of words (in case of non-default stress). At the second stage (*prosodification*), the sequence of segments is filled into the metrical frame and default stress is assigned. Note that the direction of stress processing is only specified for the prosodification stage, where rightward processing is assumed. There is no indication about the direction of stress processing at the frame generation stage. We would like to suggest that the assignment of stress to German words occurs during frame generation, given that there is no psycholinguistic evidence for a default stress pattern. At this stage, stress is computed in a leftward manner, i.e., APU stress assignment is more demanding than PU or U stress assignment. During prosodification, segments are filled into the metrical frame. According to Levelt et al. (1999), prosodification proceeds incrementally in a rightward manner. Therefore, the articulatory realization (but not the computation) of APU stress may be less demanding than the articulation of PU and U stress. This account is consistent with evidence from Italian, showing an articulatory advantage for stress positions at the left edge of a word: pseudowords stressed on the APU could be read faster than pseudowords stressed on the PU. On the other hand, the computation of main stress position for pseudowords was influenced from the phonological similarity with words on the right edge only (Burani and Arduino, 2004; Burani et al., 2013; Sulpizio et al., 2013). In other words: the processing direction between prosodification (starting at the left edge) and frame generation (starting at the right edge) may diverge. Furthermore, empirical evidence for rightward stress processing during the monitoring of lexical stress positions in Dutch reported by Schiller et al. (2006) seems to be related to the prosodification stage rather than to the frame generation stage.

Note that the lemma model is underspecified in several aspects of prosodic encoding. First, it does not incorporate the possibility of right aligning default systems (e.g., languages with fixed stress positions on the U, PU, or APU). Do the filling of the frame with segments (rightward) and the processing of default stress (leftward) occur in parallel or sequentially? Second, in some languages stress is assigned neither by default (fixed stress position) nor via lexical retrieval. Rather, it can be placed on variable positions,



which are, however, determined by fully regular rules (e.g., Cairene Arabic or Latin). At which stage of prosodic encoding do these rules operate? A third point, which is related to the second one, concerns stress assignment to non-words. How can the assignment of stress to syllables other than the default be explained? Non-words are stressed on variable positions by German participants (Röttger et al., 2012). These positions are neither restricted to the default nor fully captured by rules. If the assignment of stress involves lexical analogies – are those retrieved during frame generation? Fourth, how can different processing directions for main and secondary stress (e.g., Alber, 2005) be incorporated into the model? Obviously, many further specifications of the model have to follow in the future. Yet, the main distinction into leftward stress processing at the frame generation stage and rightward realization of stress during prosodification could already capture a number of previously contradicting findings and theories on (German) word stress.

## ACKNOWLEDGMENTS

This research was supported by a grant from the German Research Foundation (DFG grant DO 1433/1-1) and by a grant from the LOEWE initiative of excellence of the Hessian Ministry of Research and the Arts (project LingBas). The authors thank Caroline Haake for her help in data collection.

## REFERENCES

- Alber, B. (1997). “Quantity Sensitivity as the Result of Constraint Interaction,” in *Phonology in Progress – Progress in Phonology: HIL Phonology Papers*, eds Geert Booij and Jeroen van de Weijer (The Hague: Holland Academic Graphics), 1–45.
- Alber, B. (2005). Clash, lapse and directionality. *Nat. Lang. Linguist. Theory* 23, 485–542. doi: 10.1007/s11049-004-0482-1
- Benware, W. A. (1980). Zum Fremdwortakzent im Deutschen. *Z. Dialektologie Linguist.* 47, 289–312.
- Brockmole, J. R., and Logie, R. H. (2013). Age-related change in visual working memory: a study of 55,753 participants aged 8–75. *Front. Psychol.* 4:12. doi: 10.3389/fpsyg.2013.00012
- Burani, C., and Arduino, L. S. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain Lang.* 90, 318–325. doi: 10.1016/S0093-934X(03)00444-9
- Burani, C., Paizi, D., and Sulpizio, S. (2013). Stress assignment in reading Italian: friendship outweighs dominance. *Mem. Cognit.* 42, 662–675. doi: 10.3758/s13421-013-0379-5
- Dobbs, A. R., and Rule, B. G. (1989). Adult age differences in working memory. *Psychol. Aging* 4, 500–503. doi: 10.1037/0882-7974.4.4.500
- Domahs, U., Klein, E., Huber, W., and Domahs, F. (2013). Good, bad and ugly word stress – fMRI evidence for foot structure driven processing of prosodic violations. *Brain Lang.* 125, 272–282. doi: 10.1016/j.bandl.2013.02.012
- Domahs, U., Plag, I., and Carroll, R. (2014). Word stress assignment in German, English and Dutch: quantity-sensitivity and extrametricality revisited. *J. Comp. Ger. Linguist.* (in press). doi: 10.1007/s10828-014-9063-9
- Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I., and Schlewsky, M. (2008). The processing of German word stress: evidence for the prosodic hierarchy. *Phonology* 25, 1–36. doi: 10.1017/S0952675708001383
- Eisenberg, P. (1991). Syllabische Struktur und Wortakzent: Prinzipien der Prosodik deutscher Wörter. *Z. Sprachwissenschaft.* 10, 37–64. doi: 10.1515/zfs.1991.10.1.37
- Féry, C. (1986). Metrische Phonologie und Wortakzent im Deutschen. *Studium Linguistik* 20, 16–43.
- Féry, C. (1998). German word stress in optimality theory. *J. Comp. Ger. Linguist.* 2, 101–142. doi: 10.1023/A:1009883701003
- Giegerich, H. (1985). *Metrical Phonology and Phonological Structure: German and English*. Cambridge: Cambridge University Press.
- Goedemans, R., and van der Hulst, H. (2014). “Fixed stress locations,” in *The World Atlas of Language Structures Online*, eds M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology).
- Hall, T. A. (1992). *Syllable Structure and Syllable-Related Processes in German*. Tübingen: Niemeyer. doi: 10.1515/9783111657882
- Hayes, B. (1995). *Metrical Stress Theory. Principles and Case Studies*. Chicago: University of Chicago Press.
- Heisterueber, M., Klein, E., Willmes, K., Heim, S., and Domahs, F. (2014). Processing word prosody – behavioral and neuroimaging evidence for heterogeneous performance in a language with variable stress. *Front. Psychol.* 5:365. doi: 10.3389/fpsyg.2014.00365
- Janssen, U. (2003a). Stress assignment in German patients with surface dyslexia. *Brain Lang.* 87, 114–115. doi: 10.1016/S0093-934X(03)00225-6
- Janssen, U. (2003b). *Untersuchungen zum Wortakzent im Deutschen und Niederländischen*. Doctoral dissertation, University of Düsseldorf, Düsseldorf, Germany.
- Janßen, U., and Domahs, F. (2008). Going on with optimised feet: Evidence for the interaction between segmental and metrical structure in phonological encoding from a case of primary progressive aphasia. *Aphasiology* 22, 1157–1175. doi: 10.1080/02687030701820436
- Knaus, J., and Domahs, U. (2009). Experimental evidence for optimal and minimal metrical structure of German word prosody. *Lingua* 119, 1396–1413. doi: 10.1016/j.lingua.2008.04.002
- Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/S0140525X99001776
- Murre, J. M. J., Janssen, S. M. J., Rouw, R., and Meeter, M. (2013). The rise and fall of immediate and delayed memory for verbal and visuospatial information from late childhood to late adulthood. *Acta Psychol.* 142, 96–107. doi: 10.1016/j.actpsy.2012.10.005
- Röttger, T. B., Domahs, U., Grande, M., and Domahs, F. (2012). Structural factors affecting the assignment of word stress in German. *J. Ger. Linguist.* 24, 53–94. doi: 10.1017/S1470542711000262
- Schiller, N. O., Jansma, B. M., Peters, J., and Levelt, W. J. M. (2006). Monitoring metrical stress in polysyllabic words. *Lang. Cogn. Process.* 21, 112–140. doi: 10.1080/01690960400001861
- Sulpizio, S., Arduino, L. S., Paizi, D., and Burani, C. (2013). Stress assignment in reading Italian polysyllabic pseudowords. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 51–68. doi: 10.1037/a0028472
- Tappeiner, E., Domahs, U., and Domahs, F. (2007). Wortakzent im Sprachkontakt Deutsch-Italienisch. *Z. Dialektologie Linguist.* 74, 266–291.
- Vennemann, T. (1991). Skizze der deutschen Wortprosodie. *Z. Sprachwissenschaft* 10, 86–111. doi: 10.1515/zfs.1991.10.1.86
- Wiese, R. (1996). *The Phonology of German*. Oxford, UK: Oxford University Press.
- Wurzel, W. U. (1970). Der Fremdwortakzent im Deutschen. *Linguistics* 8, 87–108. doi: 10.1515/ling.1970.8.56.87
- Wurzel, W. U. (1980). Der deutsche Wortakzent: Fakten – Regeln – Prinzipien. *Ein Beitrag zu einer natürlichen Akzenttheorie. Z. Germanistik* 3, 299–318.
- Zimmermann, P., and Fimm, B. (1993). Testbatterie zur Aufmerksamkeitsprüfung. Herzogenrath: Psytest.
- Zonneveld, W., Trommelen, M., Jessen, M., Bruce, G., and Árnason, K. (1999). “Word stress in West-Germanic and North-Germanic languages,” in *Word Prosodic Systems*, ed. Van der Hulst (Berlin: Mouton), 477–603.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 January 2014; accepted: 23 May 2014; published online: 11 June 2014.  
 Citation: Domahs F, Grande M, Huber W and Domahs U (2014) The direction of word stress processing in German: evidence from a working memory paradigm. *Front. Psychol.* 5:574. doi: 10.3389/fpsyg.2014.00574  
 This article was submitted to Language Sciences, a section of the journal *Frontiers in Psychology*.  
 Copyright © 2014 Domahs, Grande, Huber and Domahs. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Developing biases

Ruben van de Vijver\* and Dinah Baer-Henney

Institut für Sprache und Information, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

## Edited by:

Ulrike Domahs, University of  
Marburg, Germany

## Reviewed by:

LouAnn Gerken, University of  
Arizona, USA

Frank Zimmerer, Universität des  
Saarlandes, Germany

## \*Correspondence:

Ruben van de Vijver, Institut für  
Sprache und Information,  
Heinrich-Heine-Universität  
Düsseldorf, Universitätsstrasse 1,  
40225 Düsseldorf, Germany  
e-mail: ruben.vijver@hhu.de

German nouns may alternate from singular to plural in two different ways. Some singular forms that end in a voiceless obstruent have a plural in which this obstruent is voiced. Another alternation concerns the vowel. Some singular forms with a back vowel have a plural form in which this back vowel is front. For each noun it has to be established individually whether it alternates or not. The voicing alternation is phonetically grounded, but the vowel alternation is not. Knowledge about such alternations involves two things. First, it involves knowledge of which words alternate and which words do not and second, it involves the ability to extend the alternations to novel words. We studied the knowledge of which words alternate and the proportion to which they alternate in two corpus studies. We studied the knowledge of speakers concerning which words alternate and what generalizations can be based upon these words by means of a production study. The production study involved words and nonces. We asked twenty 5 year-olds, twenty 7 year-olds, and ten adults to produce the plural for a given singular word and a plural for a given singular nonce. In the corpus study we found that both alternations occur with the same frequency. In the production of alternations in words we found that participants in all age groups make few mistakes. With respect to the production of alternations in nonce words, we found that the proportion of voicing alternations decreases with age, while the proportion of vowel alternations increases. We explain this change in the ability to generalize the alternations to nonces on the basis of the confidence speakers can have in a generalization. Young children have a small lexicon and they can form relatively unreliable generalizations on lexical distributions. They are, however, proficient users of language and have great phonetic experience. They can more confidently form generalizations on the basis of this experience. Adults have a large lexicon and, as a consequence, they can confidently form generalizations based on their lexicon. In addition, they know that many alternations are not based on phonetic considerations.

**Keywords:** language acquisition, morphophonology, bias, voicing alternations, vowel alternations, production test

## 1. INTRODUCTION

The pronunciation of a word often varies with morphological context. Such variation is referred to as an alternation. In this paper we will focus on two alternations in German nouns. An example of the first alternation is provided by the singular and the plural of the word [bɛʁk] *mountain* (*Berg*). The singular ends in a voiceless obstruent, but this obstruent is pronounced as voiced in the plural [bɛʁgə] *mountains* (*Berge*). This alternation is referred to in this paper as a *voicing* alternation. An example of the second alternation is provided by the singular and the plural of the word [ku:] *cow* (*Kuh*). The back vowel in the singular corresponds to a front vowel in the plural: [ky:ə] *cows* (*Kühe*). This alternation is referred to in this paper as a *vowel* alternation. Both alternations are unpredictable in the sense that one needs to know whether or not a word alternates; many words have no alternation<sup>1</sup>.

The voicing and the vowel alternation differ in their phonetic grounding. The voicing alternation has a phonetic motivation.

A voicing contrast is difficult to perceive word-finally (Steriade, 1997) and voiced obstruents are easier to produce between sonorants than voiceless obstruents (Westbury and Keating, 1986). The vowel alternation is not phonetically motivated in contemporary German. It is fossilized from a vowel harmony process that is no longer productive in German (Klein, 2000).

Native speakers have knowledge of such alternations in two different ways. The first aspect of such knowledge concerns knowledge of alternations in words. If native speakers know a singular and the corresponding plural they have knowledge of this alternation. The second aspect of this knowledge involves the ability to generalize an alternation to novel words. This latter aspect of the knowledge of a native speaker goes beyond knowing just a list of words and suggests that native speakers have knowledge of relations among words (Pierrehumbert, 2000).

In order to generalize alternations to novel words, speakers may rely on different sources of information. One important source of information is frequency in the input (Bybee, 2001a,b, 2006, 2007) and frequency in the lexicon of a speaker (Pierrehumbert, 2006). Pierrehumbert found that velar softening in English—alternations such as found in the pair [əlektʁɪk] ~

<sup>1</sup>The acquisition of suffixes—including vowel alternations—has received attention in the literature but this strand of research ignores voicing alternations (Köpcke, 1988; Clahsen et al., 1992; Kauschke et al., 2011).

[əlektrɪsɪti]—are produced rarely in nonces, and that their production depends on the knowledge of many latinate words by the participant. Buckler (2014) investigates the acquisition of voicing alternations in Dutch and German and she finds that German children are able to recognize a voicing alternation at 9 months of age, but Dutch children do not. She also counted the frequency of the voicing alternations in both languages and finds that it is more frequent in German than in Dutch. Her finding indicates that frequency is indeed an important factor in the acquisition of alternations. In order to achieve knowledge about which words alternate the learner needs to acquire a lexicon from the input—the surrounding language. Once learners have acquired a list of words they can form generalizations over these.

Another important source of information is the phonetic grounding of an alternation. If an alternation facilitates the pronunciation or the perception of a word it may be easier to generalize it to novel items than when an alternation does not improve production or perception (Hayes and Steriade, 2004; Hayes and Wilson, 2008; Baer-Henney and van de Vijver, 2012). Baer-Henney and van de Vijver study the influence of frequency and phonetic grounding on the acquisition of morphophonological alternations by Germans in using an artificial language learning experiment. Baer-Henney and van de Vijver (2012) created an artificial language in which the vowel of the plural suffix harmonized with the vowel of the stem. The suffix harmonized for the feature [back] in one language and in the other language the backness of the suffix vowel was associated with the feature [lax] of the stem vowel. The first alternation is phonetically grounded, but the second alternation is not. The alternation in each language was provided as input to the learner in two frequency conditions. In one condition the plurals were 50% of the total amount of input and in another frequency condition the plurals made up 25% of the total amount of input. This created four different artificial languages, a frequent and an infrequent backness language and a frequent and an infrequent laxness language. After a phase with only input, participants were asked to produce a plural. In the frequent condition both the backness alternation and the laxness alternation was learned. In the infrequent condition the backness alternation was generalized more often than the laxness alternation. This findings suggests that, in addition to a frequency bias, there is also a bias for phonetically based alternations. What is not known, however, is whether these biases exert an equal influence at all points in the acquisition of a language.

One reason these biases—one for frequency in the speaker's lexicon and one for the phonetic grounding of an alternation—may develop over time is that the confidence in these biases may change over time. It is reasonable to suppose that a learner relies on a bias in proportion to the amount of confidence it inspires. Generalizations in which she can place greater confidence are preferred in comparison to generalizations in which it can have less confidence (Mikheev, 1997); see also Albright and Hayes (2003) and Pierrehumbert (2006). One source of confidence in generalizations is frequency. If a particular pairing occurs very frequently and is drawn from a large pool of samples it is predicted to lead to a generalization in which it can have great confidence. If, on the other hand, such pairings come from a small sample the cognitive system can have less confidence in the generalization. There is a

difference between a pairing that occurs in 1 case out of a sample of 5 and a pairing that occurs in 100 cases out of a sample of 500 (Mikheev, 1997; Albright and Hayes, 2003; Pierrehumbert, 2006). The influence of frequency on generalizations of alternations is uncontroversial, but it is not clear whether the influence is the same for all age groups. Children have a relatively small lexicon and may therefore be skeptical about generalizations based on lexical frequency alone in comparison to adults. They are, however, experienced speakers and listeners. They can use this experience to derive generalizations. Children are, however, experienced speakers and listeners and consequently they rely more on their knowledge of speaking and listening in order to form a generalization.

We are now in a position to formulate our hypotheses. In order to learn about alternations children need to know a number of words with an alternation. We expect that, if this is the case, this is reflected in their production of such alternations in words. If this is true, we expect that children use this information to form generalizations about the alternations that they can apply to nonces. If in the input to the children the alternations are evenly distributed and this is their primary source of information for generalizations then we expect that in the production of alternations in nonces the proportions found in the input will be reflected.

If, on the other hand, the input frequency is not the sole determinant of the generalizations and that the evidence is weighted on the basis of the amount of certainty it provides, we expect that children place more confidence in the phonetic grounding of an alternation than adults. Children have a small lexicon and, therefore, place relatively little confidence in generalizations that are based on their lexicon. As they are proficient speakers and hearers they place more confidence in generalizations that reflect their knowledge of phonetics. We therefore expect that children, who have a small lexicon, may overestimate the proportion of voicing alternations—which are phonetically grounded—and underestimate the proportion of vowel alternations—which are not—both in comparison to adults.

## 2. MATERIALS AND METHODS

### 2.1. ESTIMATING THE FREQUENCY OF VOICING AND VOWEL ALTERNATIONS IN NOUNS

We first present the analysis of our corpora, since one of them served as basis for the creation of our nonces and it served as an estimate of the proportion of voicing and vowel alternations in the input of the children and adults.

We created two corpora in order to estimate the proportion of each alternation in the input. We restricted ourselves to nouns, since children appear to track frequencies per part of speech (Berko-Gleason, 1958). We counted types rather than tokens, since type frequencies is what adults track (Ernestus and Baayen, 2003).

One corpus consists of all 945 singular–plural pairs taken from a corpus based on data from the national newspaper *Frankfurter Rundschau*<sup>2</sup>. The other corpus consists of all 345 singular–plural nouns taken from the Simone-corpus, which can be found on

<sup>2</sup>The corpus was extracted with the assistance of Gerlof Bouma.

CHILDES (MacWhinney, 2000). We only used the child-directed speech from the Simone-corpus and, in addition to studying the frequency of the alternations, this phonotactics of the words in this corpus served as a basis for the phonotactics of the nonces (see Table 3).

The proportions of the type alternations are the same in both corpora (for voicing alternation: Fisher's Exact Test  $p = 0.5$ , odds ratio = 1.2, 95% confidence interval = 0.69–2.04; for vowel alternation: Fisher's Exact Test:  $p = 0.1$ , odds ratio = 1.33, 95% confidence interval = 0.93–1.89). The raw numbers are given in Tables 1, 2, in which the alternation contexts refer to the number of words ending in an obstruent for the voicing alternation and the number of words with a back stem vowel for the vowel alternation.

The frequency of both alternations in nouns is comparable. In the input children and adults are as likely to encounter a voicing alternation as a vowel alternation. This suggests that, as to the words of the children and adults, they have an equal chance of learning about voicing alternations as about vowel alternations.

## 2.2. MATERIAL

We used 24 words and created 39 nonces—phonotactically legal words that do not exist in German—in a production test (Berko-Gleason, 1958) in which we presented the participants with an item in the singular and asked them to provide the plural.

The words are common words, taken from a list of words that 2-year-olds are supposed to know (Grimm and Doil, 2000) and a few words that are part of Caroline corpus in CHILDES (MacWhinney, 2000) which Caroline used at an early age. Eight of the words ended in an obstruent, four with a voicing alternation and four without. Another eight of the words had a back vowel, four with a vowel alternation and four without. The last batch of eight words had a back vowel and a final obstruent, four of which had both a vowel and a voicing alternation and the other four had no alternation. The full list of words is given in section A.1.

To create our nonces words we extracted a corpus of 398 singular–plural pairs from a child-directed speech corpus [the Caroline corpus (MacWhinney, 2000)] and analyzed the pairs phonotactically. We wanted to ensure that the nonces resembled words, since such nonces are rated as better examples of words and are treated more like words (Frisch et al., 2000; Friedrich and

Friederici, 2005). By basing ourselves on the rhymes of a corpus of child-directed speech we ensured that the rhymes of our nonces resembled the rhymes of words used in addressing children. The distribution of environments is given in Table 3; the gap in the table concerns nonces that would not have the environment for a voicing alternation nor for a vowel alternation; they would thus fall beyond the scope of our study. The full list of nonces is given in section A.2.

## 2.3. PARTICIPANTS

We tested three groups of participants: Twenty 5-year-olds (mean age 4;9.11), twenty 7-year-olds (mean age 7;1.10), and twenty adults (mean age 29;11). They were all from the area around Potsdam, Germany, and monolingual native speakers of German.

## 2.4. PROCEDURE

The participants were seated in front of a computer. We told them that they would see pictures of familiar and unknown items on the monitor.

In a short practice session the participant was shown pictures of an apple *Apfel* [apfəl] (*Apfel*), the moon *Mond* [mo:nt] (*Mond*), a forest *Wald* [valt] (*Wald*) and, as an example of a nonce, a fantasy animal, a wug [vak]. After each picture the participant was prompted to provide the plural. If the participant did not provide any, the experimenter provided the plural. This was repeated until the participant provided the plural.

In the test phase the experimenter told the participant what was shown on the screen, for example, “Look, a [gʊəp].” (“Guck mal, ein [gʊəp].”). Then a picture with two [gʊəp]s appeared and the experimenter said “Look, now there are two! There are two...?” (“Guck mal, jetzt sind da zwei. Das sind zwei...?”) thus prompting the participant to provide the plural. Each participant was tested on 39 nonces and 24 words presented in a different, random order.

The whole session was recorded and transcribed by the experimenter and independently by the second author by auditory inspection and visual inspection. In almost all cases the raters agreed in their judgment. In the few cases where the transcribers disagreed the first author transcribed the target word blindly—without being given any information about what word or nonce

**Table 1 | Voicing alternation and vowel alternation in the Frankfurter Rundschau corpus (types).**

	Alternation contexts	Actual alternations	Proportion (%)
Voicing	270	59	21.8
Vowel	457	104	22.8

**Table 2 | Voicing alternation and vowel alternation in the child-directed speech corpus (types).**

	Alternation contexts	Actual alternations	Proportion (%)
Voicing	103	27	26.2
Vowel	234	71	30.3

**Table 3 | Phonotactics of the rhymes of the nonces.**

Voicing alternation context			
Vowel alternation context		No vowel alternation context	
Cluster	No cluster	Cluster	No cluster
Four items	Six items	Two items	Three items
[dant]	[bot]	[fɛns]	[me:k]
No voicing alternation context			
Vowel alternation context			
Cluster	No cluster		
11 items	13 items		
[dɔəm]	[mo:l]		

was intended. In all cases it could be determined what the child had said by at least two of three transcribers.

## 2.5. RESULTS

Before we discuss the results for the nonces we will briefly discuss the results for the words. The task was a production task in which the participants were free to give whatever answer as plural. In most cases the answers contained some modification of the singular we provided them with—adding a suffix or changing the vowel—but in some cases the participants did not change anything. Five year-olds produced a total of 478 responses, of which 385 (80%) had a change to the singular we provided them with and 93 (20%) words were a repetition of the singular. Seven year-olds produced 480 responses, 470 of which (97%) contained some kind of modification and 10 cases were repetitions of the singular we provided them with. Adults provided us with 479 responses, 477 (99.5%) of which contained some kind of modification and 2 (0.5%) were repetitions of the singular we provided them with. Only those responses that had any modification provide us with information about alternations, so we only included these responses in our analysis of the production of words<sup>3</sup>.

Excluding the bare nouns, the 5 year-olds produced plurals for 131 items that require voicing, such as [pfeʊt] *horse* (*Pferd*) which has the plural [pfeʊdə] (*Pferde*). They correctly produced voicing alternations in 127 words (97%) and failed to produce it in 4 words (3%). There were 126 words that end in an obstruent, but that do not require a voicing alternation, such as [fleɪk] *stain* (*Fleck*), which has as plural [fleɪkən] (*Flecken*). In 11 of these (9%) the children produced a voicing alternation. There were 126 words that require a vowel alternation in the plural, such as [ku:] *cow* (*Kuh*) which has as plural [ky:ə] (*Kühe*). In 16 of these (13%) they failed to produce a vowel alternation. There were 127 words with a back vowel in the singular that do not require a vowel alternation in the plural, such as [ʃu:] *shoe* (*Schuh*) which has as plural [ʃu:ə] (*Schuhe*). In 16 of these (13%) the children produced a vowel alternation.

As for the 7 year-olds, again excluding the bare nouns, we analyzed the errors in the way. There were 160 singulars that end in an obstruent, the plural of which has a voiced obstruent. Of these, 157 (98%) were correctly voiced and in three words (2%) they produced no voicing alternation. There were 151 words that ended in an obstruent which do not have a plural with a voicing alternation. Of these, 5 words (3%) were erroneously pronounced with a voicing alternation. There were 160 words that require a vowel alternation in the plural and in one word (0.4%) the vowel alternation was not pronounced. There were 153 words that do not require a vowel alternation, 15 of which (10%) were erroneously produced with a vowel alternation.

The adults produced all plurals correctly.

In short, as to the words, all groups of participants know the words very well. All age groups make few mistakes and the numbers of mistakes decrease with age. Their knowledge of

alternations in words provides all age groups with the basis for generalizations which can be applied to produce alternations in nonces.

Now, let us turn to the results of the nonces.

As with the words the participants sometimes repeated the nonce without any change; the plural they provided was identical to the singular we presented them with. This tendency was stronger in younger children than in adults. Five-year-olds answered with a bare stem in 538 cases (69%) and with an inflected form in 242 cases (31%). Seven-year-olds answered with a bare stem in 301 cases (39%) and with an inflected form in 477 cases (61%) and adults answered with a bare stem in 95 cases (12%) and with an inflected form in 685 cases (88%). In our analysis we included only those answers that could, in principle, be identified as a plural.

**Table 4** shows that the amount of voicing alternations produced in nonces decreases with age. Five-year-olds produced 32% voicing alternations, seven-year-olds produced 21.4% voicing alternations and adults produced 16.9% voicing alternations.

**Table 5** shows that the amount of vowel alternations increases with age. Five-year-olds produced 1.6% vowel alternations, seven-year-olds produced 5.1% vowel alternations and adults produced 10.8% vowel alternations. This is summarized in **Table 5**.

A graphical overview of these proportions of all alternations produced in nonces is shown in **Figure 1**.

We calculated the maximum likelihood of the proportion of voicing alternations for all three populations and the associated 95% confidence intervals based on a simulation of 5000 repetitions of our experiments (Gelman and Hill, 2007). We ran this analysis because the data contained too few cases to run a binomial regression analysis.

Each bell curve shows the expected distribution of the proportion of alternations for a population (Gelman and Hill, 2007). It can be seen that the distributions of the 5-year-olds and the adults do not overlap. The 5-year-olds produce more voicing alternations than the adults. The distribution of the 7-year-olds is between the distribution of the 5-year-olds and adults. This is illustrated in **Figure 2**.

The maximum likelihood distribution of the proportion of vowel alternations for all three populations and the associated 95% confidence intervals, also based on 5000 repetitions of our

**Table 4 | Nonces: voicing alternations across age groups.**

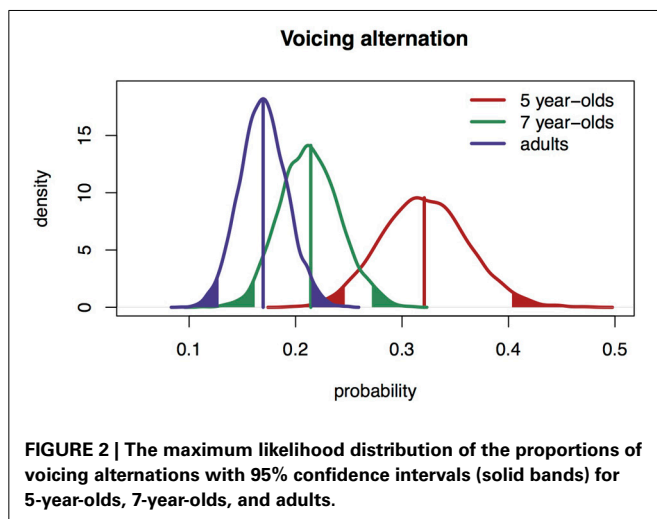
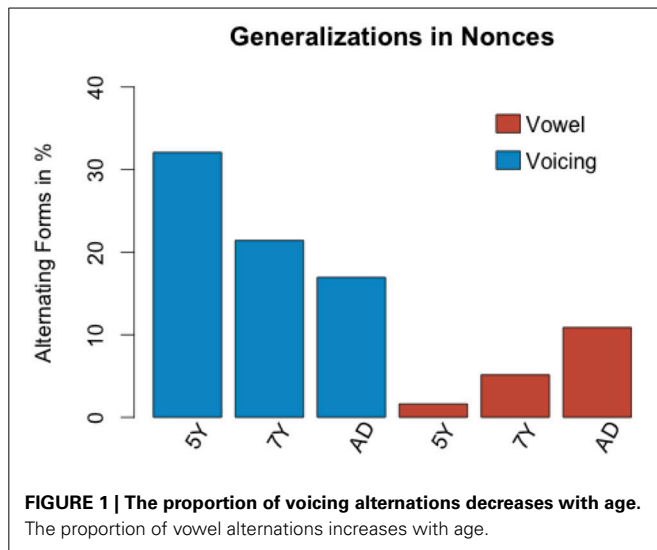
	Alternation (%)	No alternation (%)
Five-year-olds	43 (32)	91 (67.9)
Seven-year-olds	45 (21.4)	165 (78.5)
Adults	49 (12)	240 (88)

**Table 5 | Nonces: vowel alternations across age groups.**

	Alternation (%)	No alternation (%)
Five-year-olds	11 (1.6)	669 (98.7)
Seven-year-olds	35 (5.1)	645 (94.8)
Adults	74 (10.8)	606 (89.1)

<sup>3</sup>For an overview of the suffixes produced see van de Vijver and Baer-Henney (2013); in that paper there is no evidence that the suffix [s] is treated as a default. Rather its distribution is very similar to other reports in the literature (Kauschke et al., 2011) and seems to follow its lexical distribution.





experiment is shown in **Figure 3**. This distribution shows that adults produce more alternations than 5-year-olds. Seven-year-olds produce a proportion that is between the proportions of five-year-olds and adults.

### 3. DISCUSSION

We ran a production experiment in order to study the development of generalizations concerning voicing and vowel alternations in German nouns. We tested twenty 5-year-olds, twenty 7-year-olds, and twenty adults. In addition, we also studied the proportion of voicing and vowel alternations in nouns. It turns out that in two corpora the proportion of voicing and vowel alternations is the same.

Given the frequency of the alternations in nouns in the input (see **Tables 1, 2**) and both the correct productions in words and the error patterns in words one might have expected that both alternations are extended to nonces at the same rate. Both 5-year-olds and 7-year-olds produce voicing and vowel alternations in words largely correctly and, as for their errors, both groups of

participants overgeneralize the alternations to the same extent to novel words and they both fail to produce required alternations to the same extent. We found that all participant groups extended voicing and vowel alternations to nonces. With increasing age the proportion of voicing alternations falls off, while the proportion of vowel alternations climbs.

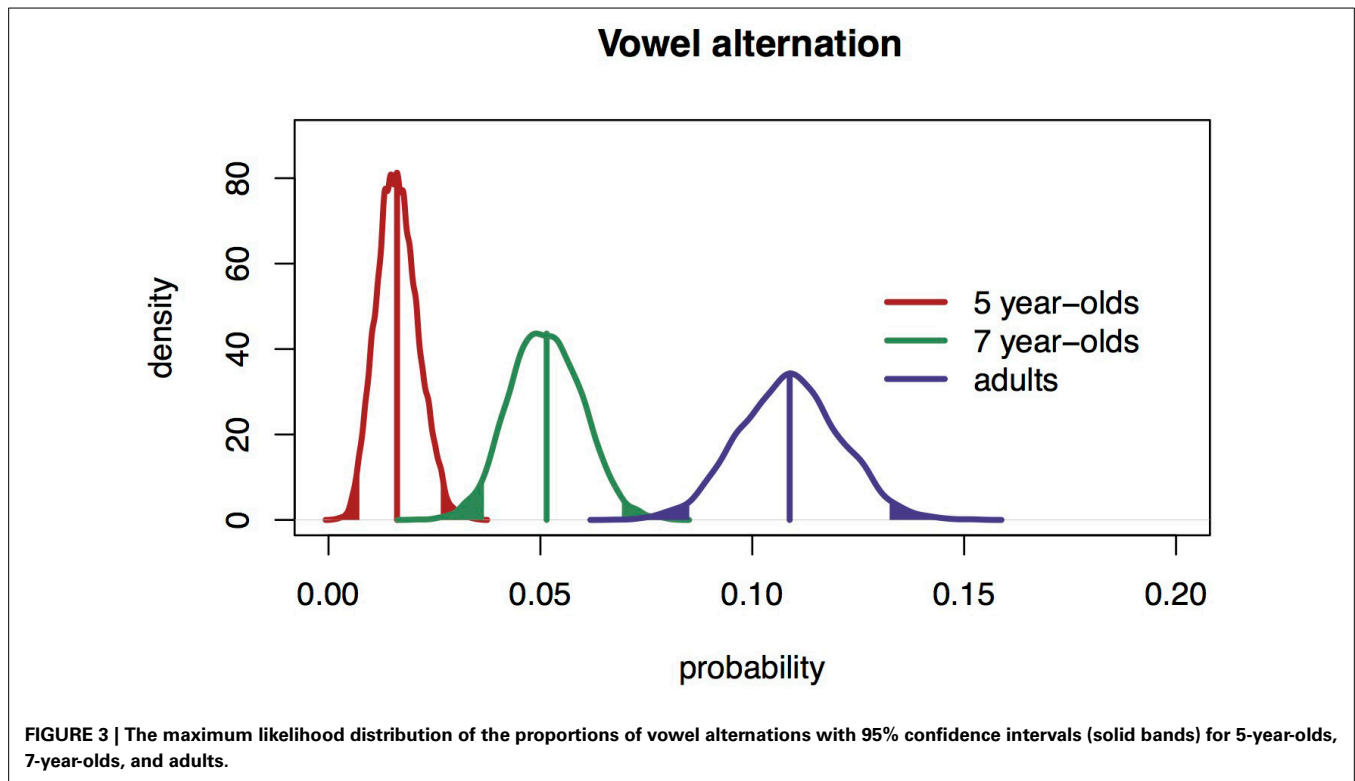
In the light of the results of our production experiment concerning words, this is unexpected. The children know both alternations in words well—which provides them with a basis for generalizations—and in their input both alternations occur with the same frequency. If the distribution of the alternations in the input is the sole source of information they use we would have expected that both are produced with the same proportion in the nonces. This is clearly not the case.

We explain this as follows. Children have evidence for pairs that alternate and pairs that do not; both in words with voicing alternations and in words with vowel alternations. It is, therefore, impossible to find a generalization which can be completely trusted. Since a 5-year-old has a relatively small lexicon any generalization based on their lexicon is necessarily based on a relatively small sample and comes, consequently, with a high degree of uncertainty. However, the 5-year-old is an experienced language user. Generalizations that are based on phonetic grounding are made with a fair amount of confidence. As a consequence, they will be more confident that a voicing alternation is warranted, as this alternation is found in the words they know, which, however, inspires little confidence, and such an alternation is phonetically grounded, which inspires much more confidence. They will have little confidence in extending vowel alternations to nonces, even though such alternations occur and in their lexicon, since they are uncertain concerning generalizations based on their lexicon and this alternation is not supported phonetically.

Seven-year-olds have a larger lexicon than five-year-olds. They have noticed that the proportion of voicing alternations is not as large as they assumed when they were five and that the proportion of vowel alternations is larger than they assumed when they were five. These insights are based on a larger sample than when they were five and, therefore, they are confident that their generalizations reflect the proportions found in their lexicon. Since their lexicon is larger it provides a more secure basis for their generalizations and they can rely less on generalization based on phonetic grounding.

Adults, of course, have the best sample of all: A large lexicon. They can be very confident that their lexicon serves as a basis for their generalizations. They can almost completely ignore any further information that derives from substance as being unreliable. This explains why in many experiments adults reflect the lexical proportions in inflections of novel words (Ernestus and Baayen, 2003), leaving only very little evidence for the presence of a bias for substantively based alternations (Albright and Hayes, 2003; Zhang and Lai, 2008, 2010; Hayes et al., 2009; Zuraw, 2010). It is interesting that the proportion of both alternations produced by adults in the nonces is similar and that the proportion of both alternations in the corpora is also the same. The fact that the absolute proportions in the data of the corpora and in the production of alternations in the nonces is different is probably a result of the fact that the phonotactics of the nonces are





a subset of the phonotactics of the words represented in the corpora.

This finding is in agreement with findings in artificial language experiments. In such experiments, where there is no lexical support, adults often show biases toward substantively based generalizations (Wilson, 2006; Finley and Badecker, 2009; Baer-Henney and van de Vijver, 2012). As their lexicon cannot serve as a secure basis for their generalizations they rely on another source of information for confidence in their generalizations: phonetic grounding.

These results can be formalized in several ways, provided the theory is able to incorporate information about frequency and is able to slightly adjust this frequency on the basis of the strength of the evidence. One formal model in which the results can be explained is the Minimal Generalized Learner proposed by Albright and Hayes (2003). In this model generalizations are the result of a comparison of two forms, for example, a singular form and a plural form. The learner takes a singular form, such as [vɛɐk] *factory, work, opus (Werk)* and its plural [vɛɐkə] (*Werke*) and compares them. In doing so, the learner concludes that forming a plural consists of adding [ə] to [vɛɐk]. The learner will encounter other pairs. For example, it will encounter the singular [bɛɐk] *mountain (Berg)* and the plural [bɛɐgə] (*Berge*). Here the learner will conclude that the first three segments [bɛɐ] remain stable over both forms and that the [k] of the singular changes to [gə] in the plural. In the case of the pair [bal] *ball (Ball)* and [bɛlə] (*Bälle*) the rule will be that the back vowel of the singular corresponds to a front vowel in the plural and that a schwa is added. In this way the learner compares all pairs it encounters and forms rules—generalizations—that map singular forms

onto plural forms. The rules themselves can be further generalized over. For example, once the learner encounters to pair [tak] *day (Tag)* and [tagə] (*Tage*) it will be able to use this rule and compare it to the rule for [bɛɐk] ~ [bɛɐgə]. The learner will notice the similarities and generalize that a singular form that ends in a dorsal voiceless stop preceded by a low vowel corresponds to a voiced dorsal stop followed by a schwa in the plural. The more pairs are captured by the rules the more confidence is placed in it and the greater the weight of the rule (Albright and Hayes, 2003). This ensures that lexical frequencies are tracked by the learner. In short, the larger the lexicon the greater the confidence in the rules. In addition the weight of the rules can be adjusted by taking into account the phonetic groundedness of a rule and giving those rules a greater weight that facilitate production or perception of the output (Wilson, 2006). The confidence placed in this additional weight is relative to the general confidence in the rules; the smaller the general confidence in the rules the larger the weight of phonetic groundedness. This interpretation agrees with experimental results on biases for phonetic groundedness (Wilson, 2006; Baer-Henney and van de Vijver, 2012) and experimental results concerning the ability to track lexical frequencies (Ernestus and Baayen, 2003). When learners have no other evidence but their knowledge of phonetics, such as in artificial language experiments, they tend to rely more on phonetic information, but if they can rely on lexical frequencies, as in nonce word productions, they will prefer that source of information.

## FUNDING

This research was supported by German Research Council Grant No. VI 223\2-1 “The acquisition of voicing and vowel

alternations in German morphophonology” to Ruben van de Vijver. The project is part of the German Research Council Priority Program 1234: “Phonological and phonetic competence: between grammar, signal processing, and neural activity.”

## ACKNOWLEDGMENTS

We would like to thank Maria Balbach, Gerlof Bouma, Hae-Eun Cho, Caroline Féry, Heiko Friehe, Henrik Froese, Susanne Genzel, Sabrina Gerth, Pauline Kortmann, Frank Kügler, Antje Sauermann, Angelina Seibt, Saskia Warzecha, the attendants of the Ph.D. Colloquium at the university of Potsdam, the attendants at the annual SPP meeting in Potsdam, Munich and Marburg. Audiences at the OCP 7 in Nice, the OCP 8 in Marrakech, and BUCLD 35, 36, and 37 in Boston. This paper has also benefitted from the input of two anonymous reviewers.

## REFERENCES

- Albright, A., and Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90, 119–161. doi: 10.1016/S0010-0277(03)00146-X
- Baer-Henney, D., and van de Vijver, R. (2012). On the role of substance, locality and amount of exposure in the acquisition of morphophonemic alternations. *Lab. Phonol.* 3, 221–256. doi: 10.1515/lp-2012-0013
- Berko-Gleason, J. (1958). The child's learning of English morphology. *Word* 14, 150–177.
- Buckler, H. M. (2014). *The Acquisition of Morphophonological Alternations Across Languages*. Ph.D. thesis, Radboud University Nijmegen.
- Bybee, J. (2001a). *Phonology and Language Use*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511612886
- Bybee, J. (2001b). *Usage-Based Phonology*. Cambridge: Cambridge University Press.
- Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language* 82, 711. doi: 10.1353/lan.2006.0186
- Bybee, J. (2007). *Frequency of Use and the Organization of Language*. Oxford, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195301571.001.0001
- Clahsen, H., Rothweiler, M., Woest, A., and Marcus, G. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition* 45, 225–255. doi: 10.1016/0010-0277(92)90018-D
- Ernestus, M., and Baayen, H. (2003). Predicting the unpredictable: interpreting neutralized segments in Dutch. *Language* 79, 5–38. doi: 10.1353/lan.2003.0076
- Finley, S., and Badecker, W. (2009). Artificial language learning and feature-based generalization. *J. Mem. Lang.* 61, 423–437. doi: 10.1016/j.jml.2009.05.002
- Friedrich, M., and Friederici, A. D. (2005). Phonotactic knowledge and lexical-semantic processing in one-year-olds: brain responses to words and non-sense words in picture contexts. *J. Cogn. Neurosci.* 17, 1785–1802. doi: 10.1162/089892905774589172
- Frisch, S., Large, N., and Pisoni, D. (2000). Perception of wordlikeness: effects of segment probability and length on the processing of nonwords. *J. Mem. Lang.* 42, 481–496. doi: 10.1006/jmla.1999.2692
- Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, NY: Cambridge University Press.
- Grimm, H., and Doil, H. (2000). *ELFRA 2-Elternfragebogen für zweijährige Kinder*. Göttingen: Hogrefe.
- Hayes, B., and Steriade, D. (2004). “Introduction: the phonetic bases of phonological markedness,” in *Phonetically Based Phonology*, eds B. Hayes, D. Steriade, and R. M. Kirchner (Cambridge, UK: Cambridge University Press), 1–33. doi: 10.1017/CBO9780511486401
- Hayes, B., and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguist. Inq.* 39, 379–440. doi: 10.1162/ling.2008.39.3.379
- Hayes, B., Zuraw, K., Siptár, P., and Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85, 822–863. doi: 10.1353/lan.0.0169
- Kauschke, C., Kurth, A., and Domahs, U. (2011). Acquisition of German noun plurals in typically developing children and children with specific language impairment. *Child Dev. Res.* 2011, 17. doi: 10.1155/2011/718925
- Klein, T. B. (2000). “Umlaut” in *Optimality Theory: A Comparative Analysis of German and Chamorro*. Tübingen: Niemeyer. doi: 10.1515/9783110915570
- Köpcke, K. (1988). Schemas in German plural formation. *Lingua* 74, 303–335. doi: 10.1016/0024-3841(88)90064-2
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk: The Database*, Vol. 2. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mikheev, A. (1997). Automatic rule induction for unknown-word guessing. *Comput. Linguist.* 23, 405–423.
- Pierrehumbert, J. (2000). What people know about sounds of language. *Stud. Linguist. Sci.* 29, 111–120.
- Pierrehumbert, J. (2006). “The statistical basis of an unnatural alternation,” in *Papers in Laboratory Phonology VIII*, eds L. Goldstein, D. Whalen, and C. Best (Berlin: Mouton de Gruyter), 81–107.
- Steriade, D. (1997). *Phonetics in Phonology: The Case of Laryngeal Neutralization*. MS., University of California at Los Angeles.
- van de Vijver, R., and Baer-Henney, D. (2013). “On the development of the productivity of suffixes,” in *Proceedings of the 37th Annual Boston University Conference on Language Development*, Vol. 37, eds S. Baiz, N. Goldman, and R. Hawkes (Somerville, MA: Cascadia Press).
- Westbury, J., and Keating, P. (1986). On the naturalness of stop consonant voicing. *J. Linguist.* 22, 145–166. doi: 10.1017/S0022226700010598
- Wilson, C. (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cogn. Sci. Multidiscip. J.* 30, 945–982. doi: 10.1207/s15516709cog0000\_89
- Zhang, J., and Lai, Y. (2008). “Phonological knowledge beyond the lexicon in Taiwanese double reduplication,” in *Interfaces in Chinese Phonology: Festschrift in Honor of Matthew Y. Chen on His 70th Birthday*, eds Y. E. Hsiao, H.-C. Hsu, L.-H. Wee, and D.-A. Ho (Taiwan: Academica Sinica), 183–222.
- Zhang, J., and Lai, Y. (2010). Testing the role of phonetic naturalness in Mandarin tone sandhi. *Phonology* 27, 153–201. doi: 10.1017/S0952675710000060
- Zuraw, K. (2010). A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Nat. Lang. Linguist. Theor.* 28, 417–472. doi: 10.1007/s11049-010-9095-z

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 January 2014; accepted: 04 June 2014; published online: 24 June 2014.  
Citation: van de Vijver R and Baer-Henney D (2014) Developing biases. *Front. Psychol.* 5:634. doi: 10.3389/fpsyg.2014.00634  
This article was submitted to *Language Sciences*, a section of the journal *Frontiers in Psychology*.  
Copyright © 2014 van de Vijver and Baer-Henney. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDIX

### A.1 WORDS

Singular	IPA	Plural	IPA	Gloss
Ball	[bal]	Bälle	[bɛlə]	ball
Baum	[baʊm]	Bäume	[bɔʏmə]	tree
Bett	[bɛt]	Betten	[bɛtən]	bed
Bild	[bɪlt]	Bilder	[bɪldə]	picture
Brot	[bro:t]	Brote	[bro:tə]	bread
Bus	[bʊs]	Busse	[bʊsə]	bus
Elefant	[e:ləfant]	Elefanten	[e:ləfantən]	elephant
Fahrrad	[fa:ra:t]	Fahrräder	[fa:re:də]	bicycle
Fleck	[flɛk]	Flecken	[flɛkən]	stain
Hand	[hant]	Hände	[hɛndə]	hand
Haus	[haus]	Häuser	[hɔʏzə]	house
Katze	[katsə]	Katzen	[katsən]	cat
Kind	[kɪnt]	Kinder	[kɪndə]	child
Kleid	[klaɪt]	Kleider	[klaɪdə]	dress
Kuh	[ku:]	Kühe	[ky:ə]	cow
Licht	[lɪçt]	Lichter	[lɪçtə]	light
Maus	[maʊs]	Mäuse	[mɔʏzə]	mouse
Ohr	[o:ə]	Ohren	[o:rən]	ear
Pferd	[pfɛʀt]	Pferde	[pfɛʀdə]	horse
Puppe	[pʊpə]	Puppen	[pʊpən]	doll
Schaf	[ʃa:f]	Schafe	[ʃa:fə]	sheep
Schiff	[ʃɪf]	Schiffe	[ʃɪfə]	boat
Schuh	[ʃu]	Schuhe	[ʃu:ə]	shoe
Stuhl	[ʃtu:l]	Stühle	[ʃty:lə]	chair

### A.2 NONCES

IPA
[bans]
[bɔt]
[dant]
[daʊl]
[dɪnt]
[dɔ:tə]
[dɔəm]
[dʊs]
[fam]
[fɛns]
[gal]
[gɔ:p]
[gu:l]
[gʊmə]
[kaʊs]
[ko:n]
[ko:pə]
[la:zə]
[lʊm]
[mare]
[me:k]
[mo:l]
[na:kə]
[nu:fə]
[paʊfə]
[paɪt]
[pɔn]
[kvasə]
[rʊlə]
[za:ə]
[ʃʊnt]
[zɔ:nə]
[zuk]
[ta:f]
[taʊlə]
[tɪs]
[tʊŋ]
[vɔkə]
[vo:t]



# Do listeners recover “deleted” final /t/ in German?

Frank Zimmerer<sup>1,2\*</sup> and Henning Reetz<sup>2</sup>

<sup>1</sup> Department of Computational Linguistics and Phonetics, Saarland University, Saarbrücken, Germany

<sup>2</sup> Institute of Empirical Linguistics, Goethe University, Frankfurt, Germany

## Edited by:

Ulrike Domahs, University of Marburg, Germany

## Reviewed by:

Pienie Zwitserlood, University of Munster, Germany

Nicolai Pharao, University of Copenhagen, Denmark

Adrian Leemann, University of Zürich, Switzerland

## \*Correspondence:

Frank Zimmerer, FR 4.7  
Computational Linguistics and  
Phonetics, Saarland University,  
Campus C72, D-66123 Saarbrücken,  
Germany  
e-mail: zimmerer@  
coli.uni-saarland.de

Reduction and deletion processes occur regularly in conversational speech. A segment that is affected by such reduction and deletion processes in many Germanic languages (e.g., Dutch, English, German) is /t/. There are similarities concerning the factors that influence the likelihood of final /t/ to get deleted, such as segmental context. However, speakers of different languages differ with respect to the acoustic cues they leave in the speech signal when they delete final /t/. German speakers usually lengthen a preceding /s/ when they delete final /t/. This article investigates to what extent German listeners are able to reconstruct /t/ when they are presented with fragments of words where final /t/ has been deleted. It aims also at investigating whether the strategies that are used by German depend on the length of /s/, and therefore whether listeners are using language-specific cues. Results of a forced-choice segment detection task suggest that listeners are able to reconstruct deleted final /t/ in about 45% of the times. The length of /s/ plays some role in the reconstruction, however, it does not explain the behavior of German listeners completely.

**Keywords:** segment reconstruction, deleted t, perception of deletion, German, natural speech processes

## INTRODUCTION

In normal conversational speech situations, speakers seem to be rather “careless.” One of the most striking results of this careless speech is that speakers often reduce words (e.g., Ernestus, 2000; Johnson, 2004; Zimmerer, 2009). Reductions are rather “minimal” when the number of segments remains unchanged. For instance, segments can be assimilated as in German *Senf* (“mustard”) which may be produced as [zɛŋf] instead of [zɛnf] (e.g., Zimmerer et al., 2009). Segments can be lenited, for example, medial /t/ may be produced as (reduced) flap in American English (e.g., Kiparsky, 1979; Patterson and Connine, 2001; Connine, 2004; Fukaya and Byrd, 2005; Tucker, 2007; Warner and Tucker, 2011 and references therein). However, reduction processes can alter the pronunciation of words more dramatically. For instance, segments can be deleted, which might even have an impact on the syllabic structure of words (e.g., Johnson, 2004; Zimmerer, 2009).

Generally, it seems that listeners are unimpressed by reductions (including deletions) in normal listening conditions despite the high amount of reductions that speakers produce. It is, however, not yet understood, how they deal with these reductions and deletions, which may be due to the use of ideal, unreduced stimuli (so called “laboratory speech”) in many perception experiments. Although there is a better control over what listeners hear in these circumstances, experiments using exclusively laboratory speech might not tell us how listeners deal with reductions occurring in natural speech. This article aims to better understand the processes that lead to the ease of perception of reduced speech.

A number of studies that used reduced items in perception experiments showed that reduced words seem to be harder to process by listeners, especially if the sentential context is not present (e.g., Pickett and Pollack, 1963; Pollack and Pickett, 1963;

Ernestus et al., 2002; Ernestus and Baayen, 2007; Zimmerer, 2009; van de Ven et al., 2011; Zimmerer et al., 2012), which appears to be somewhat contradicting the observation that listeners usually fair well in recognizing what has been said. One possible explanation for the apparent ease of perception of reduced speech is that (at least some of the) segments are reconstructed during the course of perception, due to fine phonetic detail in the input (e.g., Manuel, 1991, 1995; Hawkins, 2003; Niebuhr and Kohler, 2011). This article investigates the reconstruction of a single deleted segment, namely the alveolar voiceless stop /t/ in word-final position in German<sup>1</sup>.

Reconstruction of deleted segments has been shown for several sounds. For instance, Manuel (1991) showed that in words with a deleted Schwa (e.g., a word like *support* sounding like *sport*), fine phonetic cues were used by listeners to differentiate the two words (reduced *support* and *sport*, Manuel, 1991). Similarly, Manuel found evidence that listeners were able to reconstruct deleted /ð/ in nasal contexts, based on fine phonetic detail in synthetically created stimuli. Fine phonetic detail has also been shown to be important for the perception of more massive reductions (e.g., Hawkins, 2003; Niebuhr and Kohler, 2011).

In the literature, language-specific and cross-linguistic tendencies have been identified, both with respect to the production as

<sup>1</sup>Note, that we call the process (/t/) deletion, despite some evidence that listeners produce some phonetic cues that could be interpreted as remnants of /t/. For us, items that were transcribed as /t/-less in the corpora we used in this investigation are called deleted. This is done to differentiate this process from reductions like flapping that is more regular. Furthermore, we apply this strategy also to other cases from the literature (e.g., Manuel, 1991) which are sometimes called “seemingly” deleted.

well as the perception of deleted segments. For instance, the segment /t/ (especially in final position) is very likely to be reduced in Germanic languages (see e.g., Guy, 1980; Neu, 1980; Sumner and Samuel, 2005; Mitterer and Ernestus, 2006; Raymond et al., 2006; Zimmerer, 2009). Concerning this deletion process, researchers have identified many aspects that are very similar across a number of languages as well including aspects of the context, in which /t/ occurs (see next section). However, there are also aspects in production and perception that appear to be language-specific, for instance the role of fine phonetic detail. When speakers leave these cues to signal listeners a deletion in the speech signal, and thereby help them reconstruct segments, they arguably differ across languages, because the phonetic realization of segments (in case of /t/: the amount of aspiration, closure duration, or voicing patterns) and consequently reduction processes are quite different cross-linguistically.

### /t/ DELETION IN PRODUCTION

The voiceless coronal stop /t/ has been studied intensively. One of the reasons is that the segment is relatively frequent additionally, it is deleted quite regularly in many (Germanic) languages (e.g., Guy, 1980, 1991; Neu, 1980; Mitterer and Ernestus, 2006; Raymond et al., 2006; Zimmerer et al., 2011, 2014)<sup>2</sup>.

Cross-linguistically, (phonological) context proved to be one of the most important factors that can influence the deletion of /t/ (Mitterer and Ernestus, 2006; Raymond et al., 2006; Zimmerer et al., 2011, 2014; Mitterer and Tuinman, 2012). Mitterer and Ernestus (2006), for instance, investigated final /t/ deletion by analyzing two different corpora of spoken Dutch. These corpora differed with respect to both the speech register that was recorded, as well as the vocabulary that had been used. Their results showed that the likelihood of /t/ being deleted in Dutch was highest when preceded by the fricative /s/ and followed by bilabial sounds. Zimmerer et al. (2011, 2014) investigated the deletion of /t/ with help of two newly created corpora that were obtained with a verb paradigm production method which ensured that final /t/ was (at least in one condition) always preceded by /s/, and followed either by /s/, by /v/ or by a vowel. They found overall deletion rates of 20% in the first corpus (Zimmerer et al., 2011) and 27% (Zimmerer et al., 2014) in /-st/ contexts of the second corpus. Concerning context to the right, deletion rates were highest when the /t/ was followed by /s/ and lowest when it was followed by a vowel. Deletion rates were intermediate when final /t/ was followed by the fricative /v/. These results were stable across the two corpora. Similar results were reported by Mitterer and Tuinman (2012) who investigated final /t/ deletion by native Dutch speakers and German speakers of Dutch. The context for final /t/ was either a preceding /n/ or a preceding /s/, and /t/ deletion was most likely when preceded by /s/. However, they found different patterns for verbs and nouns in the two language groups. While Dutch speakers showed the same pattern for verbs and nouns, that is, higher

deletion rates after /s/ than after /n/, German speakers behaved (slightly) differently. Deletion rates were higher when the /t/ was preceded by /n/ for verbs produced by German speakers, but with nouns they behaved like Dutch speakers, having more deletions after /s/ than after /n/. Overall, the speakers produced deletion rates of 25% for nouns and 40% for verbs.

Across different studies, other linguistic and extra-linguistic factors have also been identified as influencing the amount of /t/ deletion, such as speaking rate (more deletion in faster speech, cf., Guy, 1980, 1991 see also Byrd and Tan, 1996), speech register (more deletions in less formal speech, cf. Mitterer and Ernestus, 2006), fluency (less deletions in non-fluent speech sections, cf., Raymond et al., 2006; Zimmerer et al., 2011, 2014), social class and dialectal differences (cf., Labov, 1966; Wolfram, 1967), speaker age (tendency for more deletions by younger speakers, cf., Guy, 1991), word category (more deletions in function words than content words, cf., Neu, 1980), relative frequency (the likelihood of /t/ to be deleted was correlated with its likelihood to be decomposed in words like *daftly* and *swiftly*, cf., Hay, 2003). Speaker gender was not consistently found to have an impact on the amount of /t/ deletion [a tendency for more deletions by male speakers compared to female speakers, was found, for instance by Wolfram, 1967; Neu, 1980, but not found by Raymond et al., 2006—analyzing the Buckeye Corpus (Pitt et al., 2007);<sup>3</sup> or Zimmerer et al., 2014].

The factors that have been found to influence /t/ deletion in different investigations of different Germanic languages seem to point to cross-linguistic similarities, such as segmental context. These similarities concern mainly factors that influence the *amount* or *likelihood* of /t/ deletion. However, the studies reported above also reveal some language-specific differences. These differences mainly concern the *actual realization* of items where the /t/ was deleted. For Dutch, for instance, Mitterer and Ernestus found that speakers kept a preceding /s/ rather short when they deleted /t/. This short /s/ could be interpreted as a cue to a deleted /t/, because in final consonant clusters, such as /st/, the /s/ was produced shorter than if it was a single segment in final position in Dutch (Mitterer and Ernestus, 2006). Interestingly, German speakers used the opposite strategy. When final /t/ in an /-st/ cluster was deleted, German speakers tended to lengthen the preceding /s/ (Zimmerer et al., 2011, 2014). The difference between Dutch and German speakers also raises interesting questions concerning the perception or restoration of deleted /t/ which will be discussed in the next section.

### /t/ DELETION IN PERCEPTION

There have been several studies addressing the perception of naturally occurring variants of /t/, including its reduction and deletion. For instance, Sumner and Samuel (2005) investigated released and glottalized variants of final /t/ and their possible role as being represented in long-term memory. When listeners encounter variants of /t/, one could argue that instead of reconstructing the segment, they could also have stored these variants directly. While in their short-term priming experiments all variants activated the correct lexical entry equally well, the results

<sup>2</sup>Regularity in this case refers to the fact that the deletions occur in noteworthy percentage of the possible cases. This does not necessarily indicate a phonological regularity. Note, however, that the notion of variable rules has been used to describe the process (e.g., Guy, 1980; see also Cedergren and Sankoff, 1974).

<sup>3</sup>They also investigated deletion of medial /d/.



from a long-term priming experiment suggested, that only canonically produced variants (with a full /t/) are stored in long-term memory. These findings indicate that variants with reduced /t/ can be handled well by listeners, but the results not lend support for a direct storage of these variants in long-term memory (Sumner and Samuel, 2005). Furthermore, even if we assume that variants of deleted /t/ were stored in memory, these items could interfere with other words that actually do not have the /t/ in their canonical form. For instance, when the German word *hau-st* (“hit—2nd PERS. SG.”) is produced without final /t/, the resulting word will be *Haus* (“house”). This means that listeners would still benefit greatly from strategies to reconstruct deleted /t/ even if (at least some of) the variants were stored in the mental lexicon.

The perception of reduced /t/ was also investigated with its flapped variant. Results were mixed: Some studies showed that flapped variants were perceived (measured as amount of lexical activation) as well as not-flapped variants in American English (Luce and McLennan, 2005; McLennan et al., 2005), other researchers showed that the reduced flap (e.g., unreduced flaps [p<sub>AF</sub>] as opposed to reduced [p<sub>AF</sub>]) was not as acceptable in perception than the unreduced flap (e.g., Tucker, 2007). A possible explanation for the difference between these studies is the status that has been assigned to the flap. Tucker assumed the flap version as canonical and only the reduced flap version as a reduced variant, McLennan and colleagues treated the flap as a reduced variant of an underlying /t/.

Concerning deleted /t/, Mitterer and Ernestus (2006) investigated the extent to which the (acoustic) patterns that were produced by speakers of Dutch—the relatively short /s/ as a cue to an underlying final consonant cluster (see /t/ deletion in production)—had an impact on perception of deleted /t/. They investigated in a perception study with resynthesized stimuli whether Dutch listeners were able to use the cues of a short /s/ to reconstruct final /t/. Furthermore, Mitterer and Ernestus were interested whether there is a difference in /t/ reconstruction depending on the context, that is, whether in the /s/ context, where /t/ deletion occurs more often, Dutch listeners are more likely to reconstruct /t/ than in the /n/ context, which is not as prone to the deletion of /t/. The findings of Mitterer and Ernestus suggest that this is indeed the case. Dutch listeners are more likely to reconstruct /t/ in the /s/ context than in the /n/ context (see also Mitterer and McQueen, 2009 for similar results), and the listeners seem to use the cue of /s/ shortness to reconstruct final /t/. This also suggests that listeners are well aware of production patterns and use this information in the perception of deleted /t/.

In another study, Mitterer and Tuinman (2012) investigated the extent to which the cues left in the acoustic signal are language-specific for the reconstruction of final /t/ with native Dutch participants and German learners of Dutch. For stimuli similar to the ones used in Mitterer and Ernestus (2006), listeners were most likely to perceive a word containing /t/ the more evidence for /t/ was present. Also, participants perceived more often a /t/, if a possible reconstruction created a word, and if the preceding context was /s/. However, there were also differences between German and Dutch listeners. German listeners were in some cases overgeneralizing in the reconstruction of /t/. Mitterer and Tuinman interpreted the difference in reconstruction rate as

partly being conditioned by transfer from their German native language, where /t/ deletion is overall less frequent, and where the cues for deleted /t/ may be different. This raises the question of the use of language-specific cues for reconstruction of /t/, which will be addressed in this article.

## RESEARCH QUESTIONS

The short overview over production and perception of deleted /t/ suggests that there are both cross-linguistic tendencies, such as context of final /t/, and language-specific processes. This article aims to investigate to what extent perception and reconstruction of /t/ deletion in German is language specific. Two research questions are addressed with a forced-choice segment detection method:

- To what extent are German listeners able to reconstruct seemingly deleted /t/? In German, about 20% of final /t/s get deleted (e.g., Zimmerer, 2009; Zimmerer et al., 2011, 2014). This means that listeners are faced with the process of /t/ deletion quite frequently.
- The second question addresses the language-specific aspect of reconstruction. Dutch listeners have been shown to use the fine phonetic details, such as /s/ length as cues for reconstruction (e.g., Mitterer and Ernestus, 2006; Mitterer et al., 2008). The findings of Zimmerer and colleagues suggest that the length cue is different in German compared to Dutch, and the results of Mitterer and Tuinman (2012) further indicate that perception might be different for German and Dutch listeners. Therefore, the question arises whether German listeners are aware of the /s/ length cue and use it to reconstruct deleted /t/.

## MATERIALS AND METHODS

### PARTICIPANTS

In all, 14 native (11 female) German speakers participated in the experiment. They were between 20 and 34 years old (mean 24.9) and were recruited at the Goethe-University, Frankfurt. They received payment in kind (cookies and coffee/tee/juice) for their participation. All of the participants spoke standard German and were born in Hesse, or Northrhine-Westphalia. All of them had lived for more than 5 years in Frankfurt area. None of the participants reported any hearing problems. They were naïve with respect to the purpose of the study, and were told about the process of /t/ reconstruction only after they completed the experiment. An experimental session lasted about 25 min, including reading the instructions and the practice session.

### MATERIALS

The basis for the experimental stimuli were utterances recorded for a verb paradigm production corpus (Zimmerer et al., 2014). The corpus consists of paradigm productions of 50 different German verb forms. Participants had produced paradigmatic cells of verbs, where a verb form was always preceded by its correct pronoun. For instance, for the verb *hauen* (“hit-INF”), a paradigmatic cell was “*du haust, sie haut, ihr haut, sie hauen*” (“you-2nd PERS SG. hit, she hits, you-2nd PERS. PL. hit, they hit”)—the context where /t/ deletion can occur is the second person singular ending of the person/number suffix—st (e.g., *du hau-st* “you 2nd PERS.

SG. hit”), which is underlined. In some cases, 3rd person singular renditions can also end in /t/, however, the morphological structure of these forms is different (e.g., *er haus-t* “he dwells”).

All items used in the forced-choice experiment were either fragments of verbs or fragments of verbs that were followed by (fragments of) pronouns from this corpus. All items were excised with Praat (Boersma and Weenink, 2013). For deleted /t/ items, fragments were extracted from the paradigms that included the vowel of the verb stem, followed by /s/, and the subsequent word could begin in either [si] (from *sie* “she/they”) or /vi/ (from *wir*—“we”)⁴. This procedure ensured that the items did not sound word-like. Therefore, possible /t/-decisions for /t/-deleted items (henceforth *Øt-items*) were not based on lexical influence, where listeners could have reconstructed a /t/ more often in order to create an existing word form in German (e.g., Ganong, 1980; Mitterer and Tuinman, 2012).

In a pretest, 93 *Øt-items* from the corpus were presented to 5 listeners for transliteration, all were students of the institute of Phonetics in Frankfurt. They were asked to write down what they heard, and had 4 s to give a transcription for every item, which they could listen to only once. The instructions were kept very simple (i.e., “write down what you think you heard”), without mentioning the segment /t/ or any reduction process or the fact that they were excised from verb paradigm productions. For the experiment reported here, 30 items which were transliterated without /t/ were used (i.e., an item counted as being transliterated by all listeners without /t/ when it was written with neither “t” nor “z”)⁵. This procedure ensured that items were used which were rated /t/-less when no close attention was paid to the segment /t/.

For the experiment, a total of 180 stimuli were excised from this verb paradigm production corpus (Table 1 gives an overview over the segmental make-up of the stimuli). Overall, 90 of these items had a /t/ present (+*t-items*). These were 45 stimuli with a /t/ (*t-items*) intervocalically, such as [a:tə], a fragment based on *braten* (“fry-INF”), and further 45 stimuli which had /t/ occurring in an /s/ context, that is, a /ts/ (which also counted as a /t/) like in [i:stsɪ] from “... *fliehst, sie* ...” (“flee 2nd PERS. SG., she”) (*ts-items*). Note that the affricate /ts/ is a phoneme of German

which is written “z” in orthography (see also footnote 5). The segment sequence [ts] in the *ts-items* could possibly also interpreted as an affricate, which could have repercussions for the results (see section Results and Discussion and Conclusions). However, these items were included in order to have [s] present in all experimental conditions and not only in the *–t* and *Øt* conditions. Furthermore, the inclusion of these items is very close to the segmental and syllabic structure of the *Øt-items*. Then there were 60 items without an underlying /t/ (*–t-items*). These were 30 stimuli which had a fricative /s/ or /z/ preceded and followed by a vowel (*s-items*) such as [i:sə] excised from *fließen* (“flow-INF”) and 30 stimuli which had another consonant (e.g., /n/), intervocalically (*n-items*) such as [anə] which was part of *bannen* (“ban-INF”). Finally, the third group (30 stimuli) had the /t/ deleted (*Øt-items*). One example is a fragment from the paradigm where in “... *fliehst, er* ...” (“flee 2nd PERS. SG., he”) the /t/ was deleted and the fragment was [i:se]. By definition, in the *Øt-items*, /t/ was deleted word finally and followed by either a consonant (/v/ or /s/), or by a vowel if the deletion lead to a sequence of two /ss/ where no boundaries between the segments could be established which were treated as one single /s/. As can be seen in Table 1, the number of +*t-items* is higher than the number of *–t-items*. Because the *Øt-items* were transcribed without “t” in the pretest, and we did not know to what extent German listeners would be reconstructing “t,” we counted these as instances of *t-less* items as well. This led to 90 items with “t” and 90 items without “t.”

The 180 items used for the experiment were produced by ten speakers of the verb paradigm production corpus mentioned above (Zimmerer et al., 2014). Individual speakers contributed between 11 and 22 items for the experiment. These speakers were also students at the Goethe-University who spoke standard German and had spent at least 5 years in Frankfurt. None of the speakers participated in the perception experiment.

## DESIGN AND PROCEDURE

An experimental trial consisted of a warning tone, which was followed by 250 ms of silence. Then, the items were presented, after which participants had 1500 ms to decide whether the item they heard had a /t/ present or not, before a new trial began. Participants were asked to press the respective response button (“t” or “no t”) on a response box with their dominant index finger. Response measurements were conducted with help of a custom-made software and hardware combination (Reetz and

⁴In a canonical standard German production, the pronoun *sie* (“she”) should be produced as [zi:]. However, in conversational speech, the initial /z/ is very often devoiced (cf. the Kiel Corpus of spontaneous speech, IPDS, 1994). Only [s] was produced in the items that were used for the experiment.

⁵In German orthography the letter “z” denotes the affricate [ts].

**Table 1 | Stimuli in the experimental conditions.**

Experimental condition	Item-type	CVC-structure	Example in IPA	Mean duration	No. of stimuli
+t	t-items	VtV	[a:tə]	262 ms	45
	ts-items	V(C)tsV	[i:stsɪ]	423 ms	45
–t	s-items	VsV	[i:sə]	320 ms	30
	n-items	VCV	[anə]	256 ms	30
Øt	Øt-items	Vst-(C)V	[i:se]	392 ms	30
SUM				332 ms	180

Kleinmann, 2003) where response boxes (with two responses buttons labeled “t” and “no t”) were connected to an external device and subsequently saved onto an Apple Mac Book Pro. Participants were tested wearing Sennheiser eH-350 headphones.

Three experimental lists were created with the 180 items. Each list was pseudo-randomized and between the experimental lists, participants could take a break. Overall, they decided 540 times whether a fragment had a /t/ present or not. Participants were tested in groups of four or less. They received written instructions before the experiments. They were asked to respond as quickly and as accurately as possible, whether the fragment they heard had a “t-sound” present or not. Participants received a training section with 14 items that were not part of the experiment. The training section preceded the actual experiment to familiarize the participants with the task and the procedure of the experiment.

Accuracy rates were calculated as percent correct responses for the respective category (after exclusion of no responses and responses that were too fast or too slow).

## PREDICTIONS

- If the /t/ was indeed completely deleted ( $\emptyset$ t-items), participants should treat these items as they would treat items where no /t/ was present ( $-t$ -items). However, based on the results reported by Mitterer and colleagues (Mitterer and Ernestus, 2006; Mitterer and Tuinman, 2012) and based on the reconstruction of other segments, listeners may be able to use the cues that are left in the speech signal to reconstruct /t/ in some cases, and we expect some qualitative difference compared to items where no /t/ is present ( $-t$ -items). We do not expect reconstruction rates that are identical to underlying items ( $-t$ - or  $+t$ -items).
- For the items where no /t/ was present or where there was an underlying /t/ (i.e.,  $-t$ -items and  $+t$ -items), we expect very high accuracy rates, because the task itself is rather straightforward and it should not lead to high error rates.

## RESULTS

In a first analysis, we investigated how accurate participants were responding to the fragments they had to listen to. Each of the participants was expected to respond to 540 fragments. Therefore, 7560 responses were expected (3 repetitions  $\times$  180 items  $\times$  14 listeners). Before examining the reconstruction of /t/, it is important to see how participants responded to clear cases where a /t/ was present or where no /t/ was present (i.e., the  $+t$ - and  $-t$ -items). The responses to  $+t$ -items and  $-t$ -items were also used to potentially exclude participants that showed a high error rate in these response categories. For the two underlying categories, 6300 responses were expected (3 repetitions  $\times$  (90 + 60) items  $\times$  14 listeners). Of these, there were 64 cases where no response was given which were excluded from further analysis. Furthermore, 117 cases of responses that were faster than 150 ms and slower than 900 ms were also excluded. Wrong responses occurred when participants pressed “t” for  $-t$ -items or “no t” for  $+t$ -items (217 responses—20  $s$ -items, 19  $n$ -items, 22  $t$ -items, 156  $ts$ -items). For the statistical analysis, errors were coded with “0” and correct responses with “1.” Table 2 gives an overview over the accuracy rates [Least Square Means (LSM) as well as Means] for the  $+t$ -

and  $-t$ -items. Accuracy rates for individual participants ranged between 99.3 and 90.7%; no participant was excluded. For the analysis, a linear mixed model was calculated with JMP (SAS, 2012), with ITEM and PARTICIPANT as random factors, item-condition ( $t$ -item,  $ts$ -item,  $s$ -item,  $n$ -item) as factor, and accuracy as dependent variable. Results indicate that ITEM-CONDITIONS differed significantly [ $F_{(3, 144.2)} = 30.47$ ;  $p < 0.0001$ ]. This was driven by the lower accuracy rate for  $ts$ -items that were significantly different from all other ITEM-CONDITIONS, which did not differ significantly, as was indicated by a Tukey HSD *post-hoc* test.

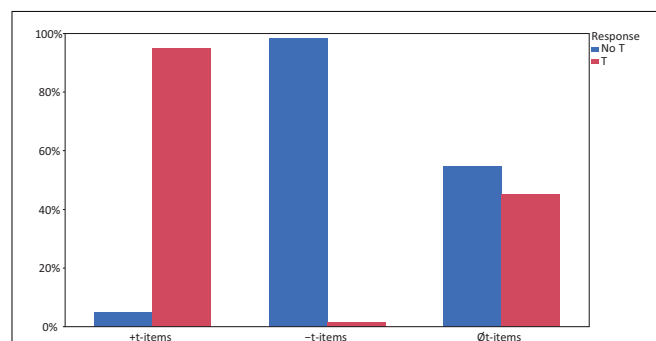
For the next analysis, that is, the comparison with  $\emptyset$ t-items, there was no difference made between  $t$ -items and  $ts$ -items (i.e., the  $+t$ -items) on the one hand and  $s$ -items and  $n$ -items (i.e., the  $-t$ -items) on the other hand.

In a next step, response categories for  $\emptyset$ t-items were compared to responses to the two underlying item categories. For this analysis, cases where no response was given (91 responses) or responses that were too fast or too slow (154 responses) were excluded from further analysis.  $\emptyset$ t-items were responded to with “t” in 45% of the time (541 responses), whereas they received a “no t” response in 655 cases (55%). For statistical analysis, we treated “no t” responses to  $\emptyset$ t-items as “correct” to be able to compare the results with the underlying items in a linear mixed model. Figure 1 shows the responses given to the respective categories ( $+t$ -,  $-t$ - and  $\emptyset$ t-items), whereas Table 3 reports the LSM for the respective categories. The linear mixed model with ITEM

**Table 2 | Least Square Means (LSM) and Means of Accuracy for the underlying items<sup>a</sup>.**

Experimental condition	Item-type	LSM accuracy	SE	Mean	SE
+t	t-items	0.988	0.009	0.988	0.004
	ts-items	0.913	0.009	0.913	0.004
-t	s-items	0.983	0.01	0.984	0.005
	n-items	0.985	0.01	0.985	0.005

<sup>a</sup>Accuracy was calculated as percentage of correct responses to the items of the respective category.



**FIGURE 1 | Percent of different responses to the respective categories, “no t” responses are in blue, while “t” responses are depicted in red.**

**Table 3 | Least Square Means (LSM) of Accuracy for different item categories.**

Experimental condition	LSM accuracy	SE
+t	0.95	0.013
−t	0.98	0.015
Øt	0.55	0.02

and PARTICIPANT as random factors, ITEM-CONDITION (+t, −t, Øt) as factor and ACCURACY (“0” for incorrect, “1” for correct) as dependent variable showed that ITEM-CONDITION was a significant factor [ $F_{(2, 176.4)} = 223.84; p < 0.0001$ ], and that each of the three conditions differed significantly. This means that deleted /t/ was reconstructed in a little less than half of the possible cases.

As indicated in the Materials and Method section, the number of stimuli was not evenly distributed for the three conditions (i.e., 90 +t-items, 60 −t-items and 30 Øt-items). This could have led participants to create a response bias. Therefore, we also used  $d'$  as a way to account for possible response biases (e.g., Stanislaw and Todorov, 1999; Macmillan and Creelman, 2005)<sup>6</sup>. The  $d'$  values were analyzed to compare sensitivity differences for the different items. A linear mixed model was calculated with PARTICIPANT as random factor and comparison as fixed factor, as well as  $d$ , as dependent variable. Results indicate that COMPARISON is a significant factor [ $F_{(2, 26)} = 74.9; p < 0.0001$ ]. A Tukey HSD test showed that all comparisons were different from each other. **Table 4** indicates the LMS for the different comparisons in this analysis. The  $d'$  analysis thus shows the same tendency as the accuracy ratio analysis.

A closer inspection of the different items showed that there was considerable variation between individual Øt-items. The amount of “no t” responses ranged from 95 to 15.4%. **Figure 2** shows the percent “no t” responses for each of the Øt-items.

In a next step, we analyzed whether participants showed different responses to the items in the course of the experiment, that is, whether they changed their reconstruction behavior over time. To this end, we calculated a linear mixed model with ITEM and PARTICIPANT as random factors, ITEM-CONDITION (+t, −t, Øt), PRESENTATION (first, second, third) and the interaction of ITEM-CONDITION and PRESENTATION as fixed factors and ACCURACY (“0” for incorrect, “1” for correct) as dependent variable. The results show that ITEM-CONDITION was a significant factor [ $F_{(2, 176.5)} = 22.94, p < 0.0001$ ], as was PRESENTATION [ $F_{(2, 7118)} = 3.82, p < 0.05$ ]. The interaction ITEM-CONDITION × PRESENTATION turned out to be significant as well [ $F_{(4, 7117)} = 8.07, p < 0.0001$ ]. In this model, concerning ITEM-CONDITION, Øt-items differed from the other two item groups, but +t-items and −t-items items did not differ significantly from each other. When we look at the factor presentation, the first PRESENTATION had an overall accuracy rate of 83.6%, the second presentation was responded to correctly in 83% of the cases, whereas in the third presentation, participants responded

**Table 4 | Least Square Means (LSM) of  $d'$  in the respective comparisons.**

Experimental condition	LSM accuracy ( $d'$ )	SE
−t vs. +t	3.98	0.17
−t vs. Øt	2.39	0.17
Øt vs. +t	1.87	0.17

correctly in 81.5% of the cases (See **Figure 3**). A Tukey HSD *post-hoc* test showed that the third condition was significantly different from the first one, but the first presentation was not different from the second, nor was the second from the third. Finally, the significant interaction of ITEM-CONDITION × PRESENTATION was driven by the fact that the Øt-items showed different accuracy rates in the three presentations. Participants responded with “no t” to Øt-items in 59.1% during the first presentation, 54.9% during the second presentation and 50.2% during the third one. The Tukey HSD *post-hoc* analysis indicates that the first and third presentations of Øt-items were different from each other, but the first and second were not, nor were the second and the third. The +t-items and the −t-items did not show significant differences in the three presentations, but were significantly different from all Øt-item presentations.

A final investigation analyzed a possible correlation between the reconstruction of /t/ and the s-length of the Øt-items. The correlation that was found was significant but not very strong [ $r_{(2)}^2 = 0.03; p < 0.0001$ ]. The analysis showed that the longer the /s/ in the signal, the more likely were participants to reconstruct /t/.

## DISCUSSION AND CONCLUSIONS

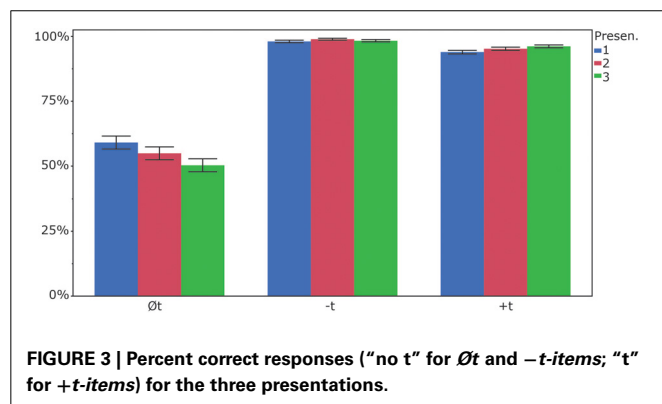
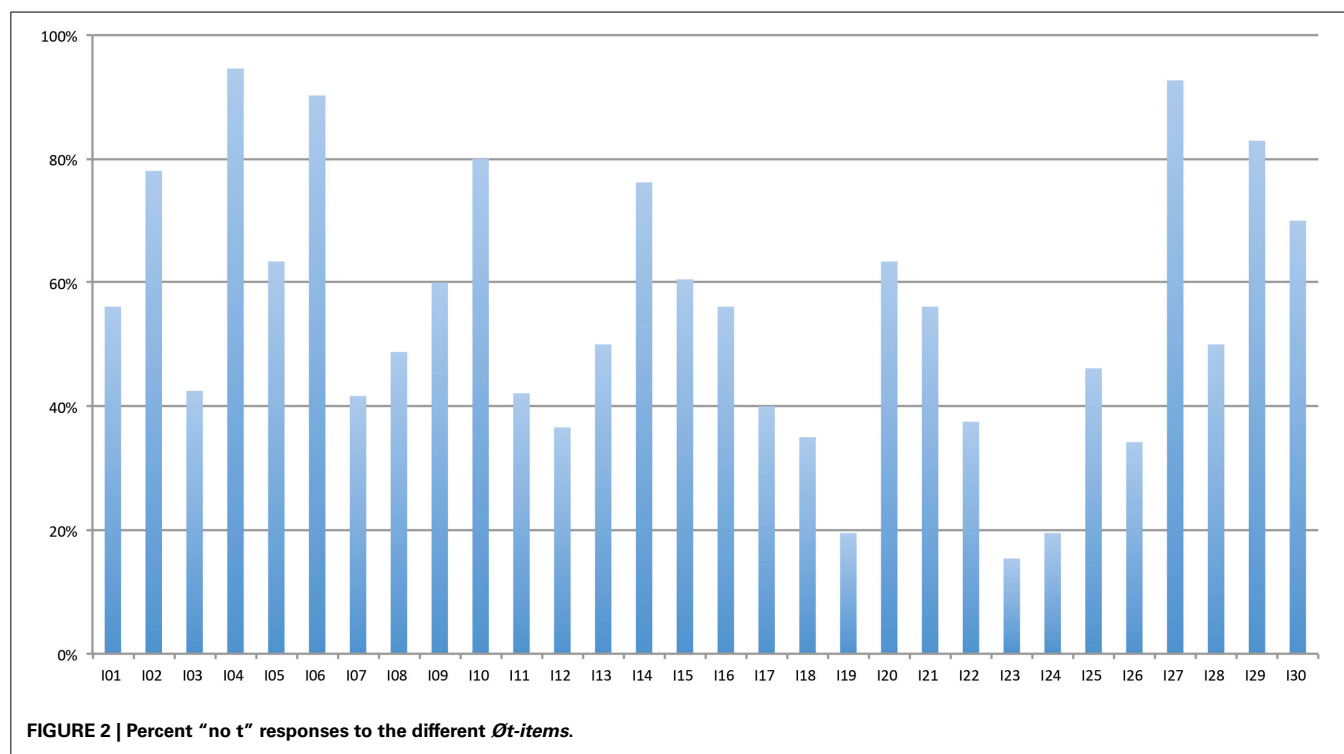
The first research question we investigated in this article concerns the extent to which German listeners are able to reconstruct seemingly deleted /t/. In German, about 20% of final /t/s get deleted (e.g., Zimmerer, 2009; Zimmerer et al., 2011, 2014). If German listeners behaved similar to Dutch listeners, reconstruction should be frequent, but it should also not occur in every instance.

The results from a forced choice phoneme detection task indicate that listeners are at least sometimes able to do reconstruct deleted /t/. When they were faced with fragments from speech with a deleted /t/, reconstruction occurred in about 45% of the cases. The amount of /t/ responses for the Øt-items is clearly different from both +t-items (with more than 95% “t” responses), as well as from −t-items (with less than 5% “t” responses). This finding is also supported by the analysis of  $d$ , where Øt-items fell in between the clear cases of −t- and +t-items.

Thus, compared to Dutch listeners, Germans seem to reconstruct deleted /t/ less often (cf. Mitterer and Ernestus, 2006; Mitterer and Tuinman, 2012). One explanation for this difference is the choice of stimuli in the experiment reported here, which are different from the experiments reported by Mitterer and colleagues. Mitterer and Ernestus used synthetically manipulated stimuli in sentential contexts (as did Mitterer and Tuinman), whereas in this study, only fragments from verbs and pronouns, that is, parts of real speech were used. These fragments were presented without context and were not word-like. Therefore, the

<sup>6</sup>In cases where participants had an accuracy rate of 100% (this occurred for −t-items only, we set the hit rate to 0.99 by adding one miss to the performance—this occurred for 7 of the participants).





stimuli in this study were on the one hand arguably closer to the kind of speech that is encountered by listeners in natural situations, since no additional manipulation (e.g., synthesis) was performed. On the other hand, they were also less natural, since usually, speech never occurs without context. Furthermore, the stimuli used here allowed for less control over what cues speakers could have left for deleted /t/ (see below). A third explanation for the different results could be that the stimuli here were not word-like, and thus, prevented listeners to adjust their perception by reconstructing a /t/. German listeners were shown to rely more on higher-level lexical knowledge when reconstructing /t/ in their foreign language Dutch (Mitterer and Tuinman, 2012). Also, in “normal” speech, German verbs are produced with a pronoun. That pronoun additionally helps to perceive the correct word and to activate the intended meaning. Therefore, this study can be seen to indirectly show further evidence for the importance of

context when listeners have to deal with deletions. If context is missing, listeners have been shown to need additional effort for successful recognition (e.g., Pickett and Pollack, 1963; Pollack and Pickett, 1963; Ernestus and Baayen, 2007; Zimmerer, 2009; van de Ven et al., 2011; Zimmerer et al., 2012). To some extent, the lower reconstruction rates of German listeners could be partly due to language-specific behavior, too. Production analyses revealed a slightly higher amount of /t/ deletion in Dutch compared to German. Therefore, Dutch listeners are faced with the deletion of /t/ more often and therefore might have a reconstruction strategy that is based more on phonetic cues (such as /s/ length or sentential context), whereas German listeners are more focused on the lexical information for reconstructing /t/.

This interpretation is also connected to the second research question we addressed in this article, concerning the impact of length of /s/ preceding a deleted /t/ on the reconstruction of /t/ (which seems to be language-specific). Dutch listeners have been shown to use the fine phonetic details of /s/ length as cues for reconstruction (e.g., Mitterer and Ernestus, 2006; Mitterer et al., 2008), but results from production analyses of German speakers indicate that German speakers behave differently from Dutch speakers, therefore, reconstruction strategies might also be different for German and Dutch listeners. The results of this study suggest that despite the consistent production of /s/ lengthening by German speakers, there seems to be only a small, but significant correlation between /s/ length and /t/ reconstruction. On the one hand, this finding can be seen as another argument in favor of language-specific reconstruction strategies of German listeners when encountering deleted final /t/. At the same time, the rather small correlation indicates that listeners are not focusing solely on /s/ length when they reconstruct /t/. The rather small effect of /s/



length could be explained partly by the nature of the stimuli (i.e., fragments without or with only minimal context). Another possible explanation is the number of stimuli with deleted /t/, which is quite small. The stimuli were all excised from the corpus without any consideration to /s/ length. Despite the consistent lengthening of /s/ in case of /t/ deletion, there is also overlap between the length of /s/ in cases where /t/ is deleted and where /t/ is not deleted (cf. Zimmerer et al., 2011, 2014). Therefore, listeners cannot be completely sure about the nature of the /s/ in the stimuli they encounter in the experiment. Furthermore, the fragments were rather short and speech rate (which may play an important role for the perception of /s/ length) cannot be estimated with high confidence by listeners, which might prevent them from using the cue of /s/ length consistently.

One result of the experiment that could lead to further research in the future concerns the differences for the underlying categories with respect to accuracy. The *ts-items* were more difficult for the participants to respond to correctly than the *t-items*, but apart from these, there was no difference between the other underlying categories. A possible explanation is that /t/ may be acoustically less salient when preceding /s/. This point is actually one of the explanations why /t/ gets deleted in such contexts in the first place. The *ts-items* were similar to *Øt-items* concerning the segmental structure, and if deletion is regarded as extreme form of reduction, maybe some of the *ts-items* were reduced to some extent. However, we also cannot rule out an orthographic influence, because in German orthography, [ts] can be written in some cases as “ts” or as “z” in others. Therefore, a transfer to the decision “t” or “no t” may be additionally difficult in these cases. Listeners could have treated these items as underlying affricates and thus refrained from responding with “t.” Especially this last possibility should be investigated in future research. It would be interesting to find out, whether listeners treat sequences of the segments [t] and [s] which are parts of the fragments across word boundaries the same or differently from underlying affricates [ts], in words like *Mütze* (“cap”) or *Herz* (“heart”), which could also be tested for the influence of orthography, because in some cases the “t” is written, in others not.

An effect that also emerged in the results was that participants responded to *Øt-items* more often with “t” in the third presentation compared to the first one. Possibly, this shows that listeners were able to learn during the experiment and focus on phonetic cues that were left in the speech signal. Despite the overall tendency to reconstruct deleted /t/ more often later in the experiment, not all items showed this trend. Individual items showed very different patterns across the three conditions some had also more “t” responses in the first presentation compared to the last one. The very different patterns of reconstruction and the finding that, overall, more deleted /t/s were reconstructed during the third presentation compared to the first one may also indicate that the overall asymmetry was not a decisive factor. If this was the case, we would have found more likely the reverse pattern. Because the number of *+t-items* is higher than *-t-items*, participants may have a “no t” bias in their response to press each button equally often. If listeners had the expectation that they should give an equal number of “t” and “no t” responses, we would expect even more “no t” responses during the third presentation, because the

asymmetry would have built up an increased response bias. This was not the case, however. Taken all presentations together, /t/ was not reconstructed in the majority of the cases by the participants.

The fact that listeners are not able to restore /t/ consistently is also shown by the varying success in /t/ restoration for individual stimuli, showing a range between 15 and 95%. No item was always (or never) restored and the variation between the items is considerable. At this point, we can only speculate about the cues that were left in the items which were leading to a more successful reconstruction compared to other cues, because /s/ length does not seem to be capturing the whole picture. Despite the significant correlation, it is extremely weak. This finding also leads to a possible extension of the experiment for further research. The choice of stimuli in this experiment was based on a pretest, where only stimuli were chosen which were never responded to with a /t/ sound. However, since reduction has been shown to be gradient, it could be interesting to include also items which were transcribed with /t/ in some of the cases in the pretest. Arguably, in these items, more cues for /t/ were left in the speech signal. The inclusion of such items could shed more light on the question what exactly the cues are and could also be used to investigate gradient activation (cf. Janse et al., 2007).

A question that additionally could arise with respect to the pretest is the difference concerning the amount of reconstruction in the pretest and in the phoneme detection experiment. At this point, we can offer only speculative explanations. First, the number of participants in the pretest was very small. Therefore, it could be possible that with more transcribers, more “t” transcriptions could have occurred for the items we chose as *Ø-items*. Furthermore, the participants for the pretest listened to each stimulus only once, and they were completely unaware of the purpose of their transcriptions and therefore did not pay attention to /t/. In the forced-choice experiment, however, participants concentrated on the presence of /t/. They had to make the decision, whether due to some phonetic detail, there could have been a /t/ present in the stimulus, and they had to do so under some time-pressure. In some cases, they arguably made a “random” decision (nearly half of the 30 *Øt-items* had accuracy rates in the range of 40–60%, cf. Figure 2). This attention to fine phonetic detail to come to a decision may also result in an overall increase in the likelihood to reconstruct /t/. And because participants were told that there is either a “t” sound present (or not), they had to decide spontaneously; there was no third option. Actually, in this respect, the rather unnatural task (“press a button for “no t” or “t”) might be even somewhat close to natural speech situation, because in these, especially, when possible ambiguities are created by the deletion of segments, listeners also have to decide (under time pressure, because conversation goes on), whether there was a /t/ present or not, and they might use all the cues they can to come to that decision (including sentential context, of course, which was absent here). At this point, we cannot be sure, which of the explanations is most accurate and some of the arguments might seem speculative, but this question is also an interesting methodological field for further research. Mitterer and colleagues circumvented many of possible problems with the stimuli by using resynthesized stimuli where it is possible to control very tightly the acoustics of what is presented. The use of non-manipulated

stimuli in our experiment does not allow for such a control. At the same time, non-manipulated stimuli might be closer to what listeners are faced with in natural occurring speech. For future research, it may be important to use both natural stimuli excised from corpora with stimuli that were artificially resynthesized.

## ACKNOWLEDGMENTS

This research was supported by the German Research Foundation (DFG), via the Priority Programme 1234 “Phonological and phonetic competence: between grammar, signal processing, and neural activity” for the project “Reductions in running speech.” We would like to thank three anonymous reviewers for their valuable comments and suggestions.

## REFERENCES

- Boersma, P., and Weenink, D. (2013). *PRAAT: Doing Phonetics by Computer*. Vers. 5.2.35. Available Online at: <http://www.praat.org/>
- Byrd, D., and Tan, C. C. (1996). Saying consonant clusters quickly. *J. Phon.* 24, 263–282. doi: 10.1006/jpho.1996.0014
- Cedergren, H. J., and Sankoff, D. (1974). Variable rules: performance as a statistical reflection of competence. *Language* 50, 333–355. doi: 10.2307/412441
- Connine, C. M. (2004). It's not what you hear but how often you hear it: in the neglected role of phonological variant frequency in auditory word recognition. *Psychon. Bull. Rev.* 11, 1084–1089. doi: 10.3758/BF03196741
- Ernestus, M. (2000). *Voice Assimilation and Segment Reduction in Casual Dutch*. Utrecht: LOT.
- Ernestus, M., and Baayen, H. (2007). “The comprehension of acoustically reduced morphologically complex words: the roles of deletion, duration, and frequency of occurrence,” in *Proceedings 16th International Congress of Phonetic Sciences (ICPhS XVI)*, Saarbrücken, 773–776.
- Ernestus, M., Baayen, H., and Schreuder, R. (2002). The recognition of reduced word forms. *Brain Lang.* 81, 162–173. doi: 10.1006/brln.2001.2514
- Fukaya, T., and Byrd, D. (2005). An articulatory examination of word-final flapping at phrase edges and interiors. *J. Int. Phon. Assoc.* 35, 45–58. doi: 10.1017/S0025100305001891
- Ganong, W. F. (1980). Phonetic categorization in auditory word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 6, 110–125. doi: 10.1037/0096-1523.6.1.110
- Guy, G. R. (1980). “Variation in the group and the individual: the case of final stop deletion,” in *Locating Language in Time and Space*, ed W. Labov (New York; London: Academic Press), 1–36.
- Guy, G. R. (1991). Contextual condition in variable lexical phonology. *Lang. Var. Change* 3, 223–229. doi: 10.1017/S0954394500000533
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *J. Phon.* 31, 373–405. doi: 10.1016/j.wocn.2003.09.006
- Hay, J. (2003). *Causes and Consequences of Word Structure*. New York, NY: Routledge.
- IPDS. (1994). *The Kiel Corpus of Spontaneous Speech*. Kiel: IPDS.
- Janse, E., Nootboom, S. G., and Quené, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Lang. Cogn. Process.* 22, 161–200. doi: 10.1080/01690960500371024
- Johnson, K. (2004). “Massive reduction in conversational American English,” in *Proceedings Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, Tokyo, 29–54.
- Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguist. Inq.* 10, 421–441.
- Labov, W. (1966). “Some sources of reading problems for Negro speakers of non-standard English,” in *New Directions in Elementary English*, ed A. Frazier (Champaign, IL: National Council of Teachers of English), 1–38.
- Luce, P. A., and McLennan, C. T. (2005). “Spoken word recognition: the challenge of variation,” in *The Handbook of Speech Perception*, eds D. B. Pisoni and E. R. Remez (Malden, MA: Blackwell), 591–609.
- Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory - A User's Guide*. 2nd Edn. Mahwah, NJ; London: Lawrence Erlbaum Associates.
- Manuel, S. Y. (1991). “Recovery of “Deleted” Schwa,” in *Proceedings Phonetic Experimental Research at the Institute of Linguistics University of Stockholm (PERILUS XIV)*, Stockholm, 115–118.
- Manuel, S. Y. (1995). Speakers nasalize /ð/ after /n/, but listeners still hear /ð/. *J. Phon.* 23, 453–476. doi: 10.1006/jpho.1995.0033
- McLennan, C. T., Luce, P. A., and Charles-Luce, J. (2005). Representation of lexical form: evidence from studies of sublexical ambiguity. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 1308–1314. doi: 10.1037/0096-1523.31.6.1308
- Mitterer, H., and Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: evidence from /t/-lenition in Dutch. *J. Phon.* 34, 73–103. doi: 10.1016/j.wocn.2005.03.003
- Mitterer, H., and McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 244–263. doi: 10.1037/a0012730
- Mitterer, H., and Tuinman, A. (2012). The role of native-language knowledge in the perception of casual speech in a second language. *Front. Psychol.* 3:249. doi: 10.3389/fpsyg.2012.00249
- Mitterer, H., Yoneyama, K., and Ernestus, M. (2008). How we hear what is hardly there: Mechanisms underlying compensation for /t/-reduction in speech comprehension. *J. Mem. Lang.* 59, 133–152. doi: 10.1016/j.jml.2008.02.004
- Neu, H. (1980). “Ranking of constraints on /t, d/ deletion in American English: a statistical analysis,” in *Locating Language in Time and Space*, ed W. Labov (New York, NY: Academic Press), 37–54.
- Niebuhr, O., and Kohler, K. J. (2011). Perception of phonetic detail in the identification of highly reduced words. *J. Phon.* 39, 319–329. doi: 10.1016/j.wocn.2010.12.003
- Patterson, D., and Connine, C. M. (2001). A corpus analysis of variant frequency in American English flap production. *Phonetica* 58, 254–275. doi: 10.1159/000046178
- Pickett, J. M., and Pollack, I. (1963). Intelligibility of excerpts from fluent speech: effects of rate of utterance and duration of excerpt. *Lang. Speech* 6, 151–164.
- Pitt, M. A., Dille, L. C., Johnson, K., Kiesling, S., Raymond, W., Hume, E., et al. (2007). *Buckeye Corpus of Conversational Speech (Final Release)*. Columbus, OH: Ohio State University (Distributor).
- Pollack, I., and Pickett, J. M. (1963). The intelligibility of excerpts from conversation. *Lang. Speech* 6, 165–171.
- Raymond, W. D., Dautricourt, R., and Hume, E. (2006). Word internal /t, d/ deletion in spontaneous speech: modelling the effects of extra-linguistic, lexical, and phonological factors. *Lang. Var. Change* 18, 55–97. doi: 10.1017/S0954394506060042
- Reetz, H., and Kleinmann, A. (2003). “Multi-subject hardware for experiment control and precise reaction time measurement,” in *Proceedings 15th International Congress of Phonetic Sciences (ICPhS XV)*, Barcelona, 1631–1634.
- SAS. (2012). *JMP*. Cary, NC: SAS Institute.
- Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* 31, 137–149. doi: 10.3758/BF03207704
- Sumner, M., and Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *J. Mem. Lang.* 52, 322–338. doi: 10.1016/j.jml.2004.11.004
- Tucker, B. V. (2007). *Spoken Word Recognition of the Reduced American English flap*. Ph. D. Thesis, University of Arizona, Tucson, AZ.
- van de Ven, M., Tucker, B. V., and Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Mem. Cogn.* 29, 1301–1316. doi: 10.3758/s13421-011-0103-2
- Warner, N., and Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *J. Acoust. Soc. Am.* 130, 1606–1617. doi: 10.1121/1.3621306
- Wolfram, W. A. (1967). *A sociolinguistic Description of Detroit Negro Speech*. Washington, DC: Center for Applied Linguistics.
- Zimmerer, F. (2009). *Reduction in Natural Speech*. Ph. D. Thesis, Goethe Universität, Frankfurt.
- Zimmerer, F., Cornell, S., Reetz, H., and Eulitz, C. (2012). Massive reduction and lexical activation. *Proc. Lab. Phonol.* 13, 153–154.
- Zimmerer, F., Reetz, H., and Lahiri, A. (2009). Place assimilation across words in running speech: corpus analysis and perception. *J. Acoust. Soc. Am.* 125, 2307–2322. doi: 10.1121/1.3021438

Zimmerer, F., Scharinger, M., and Reetz, H. (2011). When BEAT becomes HOUSE: factors of word-final /t/-deletion in German. *Speech Commun.* 53, 941–954. doi: 10.1016/j.specom.2011.03.006

Zimmerer, F., Scharinger, M., and Reetz, H. (2014). Phonological and morphological constraints on German /t/-deletions. *J. Phon.* 45, 64–75. doi: 10.1016/j.wocn.2014.03.006

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 January 2014; accepted: 24 June 2014; published online: 17 July 2014.

Citation: Zimmerer F and Reetz H (2014) Do listeners recover “deleted” final /t/ in German? *Front. Psychol.* 5:735. doi: 10.3389/fpsyg.2014.00735

This article was submitted to Language Sciences, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Zimmerer and Reetz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# An analysis of post-vocalic /s-ʃ/ neutralization in Augsburg German: evidence for a gradient sound change

Véronique Bukmaier\*, Jonathan Harrington and Felicitas Kleber

Institute of Phonetics and Speech Processing, Ludwig-Maximilians-Universität München, Munich, Germany

## Edited by:

Ulrike Domahs, University of  
Marburg, Germany, Germany

## Reviewed by:

Alexander Werth,  
Forschungszentrum Deutscher  
Sprachatlas, Germany  
Timo B. Roettger, University of  
Cologne, Germany

## \*Correspondence:

Véronique Bukmaier, Institute of  
Phonetics and Speech Processing,  
Ludwig-Maximilians-Universität  
München, Schellingstraße 3, 80799  
Munich, Germany  
e-mail: bukmaier@  
phonetik.uni-muenchen.de

The study is concerned with a sound change in progress by which a post-vocalic, pre-consonantal /s-ʃ/ contrast in the standard variety of German (SG) in words such as *west/wäscht* (/vɛst/~vɛʃt/, *west/washes*) is influencing the Augsburg German (AG) variety in which they have been hitherto neutralized as /vɛʃt/. Two of the main issues to be considered are whether the change is necessarily categorical; and the extent to which the change affects both speech production and perception equally. For the production experiment, younger and older AG and SG speakers merged syllables of hypothetical town names to create a blend at the potential neutralization site. These results showed a trend for a progressively greater /s-ʃ/ differentiation in the order older AG, younger AG, and SG speakers. For the perception experiment, forced-choice responses were obtained from the same subjects who had participated in the production experiment to a 16-step /s-ʃ/ continuum that was embedded into two contexts: /mɪst-mɪʃt/ in which /s-ʃ/ are neutralized in AG and /və'mɪsə-/və'mɪʃə/ in which they are not. The results from both experiments are indicative of a sound change in progress such that the neutralization is being undone under the influence of SG, but in such a way that there is a gradual shift between categories. The closer approximation of the groups on perception suggests that the sound change may be more advanced on this modality than in production. Overall, the findings are consistent with the idea that phonological contrasts are experience-based, i.e., a continuous function of the extent to which a subject is exposed to, and makes use of, the distinction and are thus compatible with exemplar models of speech.

**Keywords:** neutralization, sound change, dialect leveling, categorical vs. continuous, exemplar theory

## INTRODUCTION

The present study forms part of a series of investigations (e.g., Kleber, 2011; Müller et al., 2011; Harrington et al., 2012) into dialect leveling in High German varieties under the influence of Standard German (SG). Our particular concern is not just with phonological categorical changes in the direction of SG but more specifically with how such categorical changes are related to the continuously gradient variation in speech production and perception across generations of speakers. The present investigation deals with the association between the post-vocalic /s-ʃ/ contrast before /t/ in SG (e.g., *West/wäscht*; /vɛst/~vɛʃt/, engl. *west/washes*) and the Augsburg variety of German (AG) in which, at least for older, but possibly not for younger speakers, the distinction is collapsed such that these minimal pairs are neutralized as a post-alveolar fricative (i.e., /vɛʃt/ for both *West* and *wäscht*). By Augsburg variety we mean a regional variety of Standard German, which is mainly influenced by the Swabian dialect.

In Standard German, the contemporary /s-ʃ/-contrast emerged as a consequence of various sound changes. Old High German (OHG) did not distinguish between those two places of articulation for fricatives, but only had alveolar sibilants, which were realized either voiceless (fortis, /s/) or voiced (lenis, /z/). The OHG /z/ later changed into the contemporary Standard German /ʃ/ (Renn and König, 2009). In addition, /s/ shifted to

/ʃ/ in some /s+consonant/-clusters (/sC/ hereafter) from Middle High German (MHG) to SG. The shift from MHG /s/ to SG /ʃ/ took place only in syllable initial clusters (e.g., MHG *slagen* /slagən/ > SG *schlagen* /ʃlagən/, *to beat*), while in Southern German varieties this change also occurred in post-vocalic clusters (e.g., *fast*, engl. *almost*, which is /fast/ in SG but /faʃt/ in the south-west German variety of Swabian). However, while Bavarian (spoken in south-east Germany) nowadays contrasts /s/ and /ʃ/ before consonants just like SG, Swabian retains the pronunciation of /sC/-clusters as /ʃC/—not just in the deep dialect but also in the Swabian-colored, regional variety of Standard German. Thus, the Standard German phonemic contrast between post-vocalic, pre-consonantal /s/ and /ʃ/ is neutralized in favor of the post-alveolar pronunciation in Swabian, i.e., the minimal pair *West* (/vɛst/, *west*) and *wäscht* (/vɛʃt/, *washes*) are homophones when produced by a Swabian speaker. Nonetheless, in the Swabian variety the contrast between /s/ and /ʃ/ is maintained in intervocalic position (e.g., *Tasse* /tasə/, *cup*—*Tasche* /taʃə/, *bag*).

The data for the present study is taken from Augsburg—a city in Bavaria around 80 km north-west from Munich. Augsburg is situated in a transitional zone between the Bavarian and Swabian dialect areas and as a consequence, this variety has both Bavarian as well as Swabian dialect features (Nübling, 1988). In

an investigation that forms the background to the present study, Bukmaier (2010) carried out an auditory analysis to determine whether the Augsburg variety should be classified as a Swabian or a Bavarian dialect based on the proportion of Bavarian and Swabian dialect features in Augsburg speakers' productions; in order to do so, she investigated the usage of dialectal features by younger (aged 20–30 years) and older (aged 40–70 years) Augsburg speakers. Her analysis showed that AG was predominantly Swabian but that there was nevertheless a tendency for younger speakers to make greater usage of SG features. It is this latter finding that is the primary motivation for the present study that focuses on the neutralization of pre-consonantal, post-vocalic /s-/ in Augsburg German.

The phonological process of neutralization is traditionally conceived as involving a categorical change from one category to another. Nevertheless, acoustic analyses have repeatedly shown that neutralization is incomplete (Port and O'Dell, 1985; Kleber et al., 2010). Similarly, the outcome of historical sound changes is usually categorical, although there is increasing evidence that a diachronic change comes about through a gradual change from one category to another across generations (e.g., Harrington et al., 2012). Since Labov's (1963) pioneering work in sociolinguistics, so-called sound changes in progress are inferred by comparing phonetic differences across two generations of the same speech community and most often within sounds that differ in continuous acoustic parameters (as the many studies on vocalic change show, e.g., Hawkins and Midlgey, 2005) since the gradual changes are perceptible and thus more obvious. There are, however, categorical sound changes such as metathesis that are typically considered to involve no such gradual change. Similarly, the auditory analysis of the data in Bukmaier (2010) points to a categorical change amongst younger speakers from AG /f/ in clusters toward SG /s/.

On the other hand, research on assimilatory processes, in particular in /s#f/ or /f#s/ across word boundaries, has shown that sibilants vary gradually between the two places of articulation depending on the degree of assimilation (Niebuhr et al., 2008; Pouplier et al., 2011), although these fine phonetic differences may not be perceptible (Niebuhr and Meunier, 2011). Similarly, physiological studies of speech errors present evidence for gradual shifts between categories that may be perceived as clear instances of one category and may even result in auditory transcription errors (e.g., Pouplier and Hardcastle, 2005; Goldstein et al., 2007). In the light of this synchronic evidence, it seems quite possible that even these supposedly categorical diachronic changes may in fact be continuous. Thus, one of the main issues we address in this paper is whether the unmerging of /ft/ toward /st/ or /ft/ is a categorical or continuous process. A categorical change might occur lexically such that there is a discrete change for younger but not older AG speakers from /ft/ to /st/ in words such as West (SG /vest/). In a continuous change, speakers might gradually shift their production in such words between post-alveolar and alveolar productions with a greater shift toward /s/ in younger speakers.

Another major concern in this paper is whether the change affects the modalities of speech perception and production in equal measure. The arguments for parity between speech production and perception have been made across different kinds

of models including at the level of gestures (e.g., Fowler et al., 2003) and also in terms of exemplar theory (Pierrehumbert, 2002) in which speech production draws upon the same sets of exemplars that have been stored in the acoustic/auditory space of the listener's mental lexicon as a result of speech perception. With respect to some sound changes, such parity can be observed within but not between generations. An example for such a sound change in progress in which there is parity between the two modalities within a generation is the age-graded neutralization of the voicing contrast of intervocalic consonants toward the lenis variant of East Franconian speakers (Müller et al., 2011). Older East Franconians neutralize the voicing contrast of Standard German plosives in perception as well as in production, while younger East Franconians neutralize this contrast equally in production as well as in perception to a lesser extent. Nevertheless, younger East Franconians do not yet maintain the voicing contrast to the same extent as Standard German speakers. The exemplar theory not only accounts for this parity but also for the shift toward the Standard German contrast<sup>1</sup>: the more a speaker is exposed to Standard German, the more standard forms (with all the fine phonetic detail inherent to them) are added to the edge of an exemplar cloud (i.e., the density distribution of a set of exemplars across the acoustic/auditory space that constitute a phonological category) which eventually shifts in the acoustic/auditory space and then in turn causes the speakers to select more standard-like variants from the cloud for production. On the assumption that the contact with the standard variety increases with each generation of German dialect speakers, we therefore predict with respect to the present study that younger Augsburg speakers produce sibilants before /t/ in a more standard-like way than do older speakers.

At a particular point in time during the period of change, on the other hand, sound change may also present an exceptional case in which the two modalities are out of alignment with each other (Kleber et al., 2012). According to Ohala (1981, 1993), sound change is initiated by listeners' misperceptions of speakers' production. Given the vast amount of synchronic variation in speech signals (Hawkins, 2003), misperceptions may occur under certain conditions, although these misperceptions only rarely turn into a diachronic change. A similar line of argument is found in Browman and Goldstein (1991) who present evidence for articulatory gestures that overlap to such an extent that only one gesture is decoded correctly by the listener. These forms of overlap cause at first perceptual synchronic elision, which can under certain conditions result in diachronic elision. In both models it is the mismatch between production and perception that leads to sound changes on the listener's side. Applied to the present data, AG subjects might initially unmerge /ft/ as /st, ft/ in perception with production showing a greater degree of neutralization (cf. also Labov et al., 1991).

Sound changes triggered by misperceptions of or undercompensating for synchronic variation (Harrington et al., 2008; Kleber et al., 2012) are thus driven by internal or phonetic factors.

<sup>1</sup>The direction of this change is not easily accounted for by other phonetic models of sound change as phonetically lenition is much more likely to occur than fortition as the many diachronic lenitions in Romance languages show.



External or sociolinguistic factors such as social status or the prestige of a dialect (Kerswill, 2003; Labov, 2007) may, however, also play a role in diachronic changes—in particular those that are due to *dialect leveling*, which refers to the reduction of dialectal forms, as for example the increasing monophthongization of regional /ɪə/ as /e:/ in British English with the latter having a wider geographical distribution (Kerswill, 2003). The question arises whether sound changes that are triggered by sociolinguistic factors occur passively as a result of accommodation (e.g., Trudgill, 2004) or whether the speaker takes up a more active part. The model of sound change described in Lindblom et al. (1995) emphasizes the role of the speaker to a greater extent than the above-mentioned models, as it is the speakers who adapt to listeners' needs when producing speech along a continuum from hypo- to hyper-articulated speech. Sound changes may then evolve when listeners' attention is in such circumstances exceptionally directed to a word's form (i.e., its pronunciation) instead of its meaning. Perhaps speakers of regional varieties have a propensity to evaluate the word's form when they are in contact with speakers from other varieties.

The aim of the present study was to investigate whether or not Augsburg speakers completely neutralize the /s-/contrast in the production and perception of /sC/-clusters and whether the degree of neutralization is age-related in this variety, with younger Augsburg speakers tending to a more standard-like pronunciation. The analysis in this paper draws upon the classic technique of an apparent time investigation in which sound change is inferred by comparing phonetic differences across two generations. However, in contrast to almost all sociolinguistic investigations, the present study is based both on production and on the same speakers' responses to perceptual stimuli (see also Harrington et al., 2012, 2013; Kleber et al., 2012). The hypotheses for the two experiments can be formulated as follows:

H1: Augsburg speakers differentiate the /s-/contrast in /st/-clusters to a lesser extent in production than Standard German speakers.

H2: Older Augsburg speakers show a greater tendency toward neutralization of the /s-/contrast in the production of /st/-clusters than younger Augsburg speakers.

H3: Augsburg listeners differentiate the /s-/contrast in /st/-cluster to a lesser extent in perception than Standard German speakers.

H4: Older Augsburg listeners show a greater tendency toward neutralization of the /s-/contrast in the perception of /st/-clusters than younger Augsburg speakers.

## PRODUCTION EXPERIMENT

### METHODS

#### Participants

The production experiment was conducted with three different subject groups: older Augsburg speakers, younger Augsburg speakers and Standard speakers. The first group—the experimental group—contained 26 speakers of Swabian from the city of Augsburg. Eleven of these subjects were aged between 40 and 70 years (3 male and 8 female) and assigned to the older age group. 15 participants were aged between 20 and 30 years (8 male and 7 female) and assigned to the younger age group. All participants

were born/or have spent most of their lives in Augsburg. At the time of participation in this experiment all Augsburg subjects were living in Augsburg.

The second group served as a control group and included 16 Standard German-speaking subjects (two male and 14 female) aged between 20 and 30 years. The participants in this group were all either from Northern Germany or from Munich<sup>2</sup>. None of the 45 subjects reported any hearing, eye-sight, or reading problems.

Prior to the experiment the Augsburg participants were asked to fill out a questionnaire with questions about the participants' education, the length of time that they had been living in Augsburg, and a self-assessment of how much and how often they speak dialect. The AG participants were chosen in accordance to the time they had been living in Augsburg; so all the young AG subjects were living in Augsburg all of their lives and the older AG participants were living in Augsburg most of their lives (30 years and more).

The subjects of the older and the younger experimental group were tested in a quiet room at their homes. The subjects of the control group were tested in a quiet room at the university. It is possible that the difference of whether the speakers were recorded at home or not could have had an influence on the results such that those recorded at home hypoarticulated more than those in the laboratory due to the slightly more informal recording setting at home. However, we found no evidence for this from our auditory impressions of the data.

### Materials

In order to elicit productions of /st/-clusters, we designed a blending task (see also Kleber et al., 2010) in which the subjects had to combine the first syllable of one nonword with the second syllable of another nonword (see **Table 1**) in order to produce a real German word, e.g., the speaker's task was to produce the blend *Kiste* (/kɪstə/, *box*) from the two nonsense words *Küssingen* and *Wirte*.

With the exception of /u:/ in *Schuster*, the vowels /ɪ/, /ɛ/, and /ʏ/ in the initial syllables of the resulting blends were always phonologically short, which was triggered by a word medial orthographic double consonant in the first word, e.g., <ss> in *Lüssingen* (this orthographic representation corresponds to

<sup>2</sup>The dialect spoken in Munich is not affected by the dialect feature in this study, i.e., the Munich variety has exactly same /s-/ contrast distribution as Standard German.

**Table 1 | Nonwords and resulting blends.**

Word 1	Word 2	Blend
<u>K</u> üssingen (/kʏsɪŋən/)	Wir <u>t</u> e (/vɪrtə/)	<u>K</u> ü <u>st</u> e (/kʏstə/, <i>coast</i> )
<u>K</u> üssingen (/kʏsɪŋən/)	Wü <u>r</u> te (/vʏrtə/)	<u>K</u> is <u>t</u> e (/kɪstə/, <i>box</i> )
<u>L</u> üssingen (/lysɪŋən/)	Kir <u>t</u> e (/kɪrtə/)	<u>L</u> ü <u>st</u> e (/lʏstə/, pl. <i>desire</i> )
<u>L</u> üssingen (/lysɪŋən/)	Kür <u>t</u> e (/kʏrtə/)	<u>L</u> is <u>t</u> e (/lɪstə/, <i>list</i> )
Sch <u>u</u> ssingen (/ʃʊsɪŋən/)	Kir <u>t</u> e (/kɪrtə/)	Sch <u>u</u> st <u>er</u> (/ʃu:stə/, <i>cobbler</i> )
Schw <u>e</u> ssingen (/ʃvɛsɪŋən/)	Kür <u>t</u> e (/kʏrtə/)	Schw <u>e</u> st <u>er</u> (/ʃvɛstə/, <i>sister</i> )

The syllables that were blended are underlined.

the Standard German norm indicating phonemic short vowels). While the onset consonant varied, the coda consonant of the first syllable was always /s/. The final syllable of the second word was either /tə/ or /tɐ/ (see **Table 1**). The 16 filler words were disyllabic German words which did not contain any sibilants and which varied in the vowel as well as in the coda consonant of the first syllable (while the second syllable was always *-te* /tə/), e.g., *Wirte*, *Worte*, *Bunte*, *Kalte*.

In addition to the cluster blends, we obtained prototypical /s/ and /ʃ/ in intervocalic or post-vocalic position, i.e., in a non-neutralizing context in both varieties. For this purpose, subjects read aloud the following four German real words: *Biss* (/bɪs/, *bite*), *wisse* (/vɪsə/, *to know*), *Busch* (/bʊʃ/, *bush*), and *Tusche* (/tʊʃə/, *India ink*). In order to minimize any coarticulatory effects, /s/ and /ʃ/ were combined with /ɪ/ and /ʊ/, respectively.

### Experimental set-up, digitization, labeling

The recordings were made with the *SpeechRecorder* software (version 2.6.14; see Draxler and Jänsch, 2004), an audio interface (M-Audio Fast Track) and a stereo headset (Beyer dynamics). Each of the six target blends together with eleven distractor blends were repeated ten times and presented in randomized order on a MacBookPro computer screen (in total 170 tokens). Following the blending task, but within the same session and experimental set-up, the subjects were presented with three repetitions of each of the German real words (in total 12 tokens). In both tasks, the subjects had to produce each word within a time slot of 1 s, which was then followed by an automatic pause of 0.8 ms before the next item was presented. In total, each subject produced 182 words.

The words were digitized at 44.1 kHz. All of the data were segmented and labeled automatically into phonetic segments using the Munich Automatic Segmentation System (MAuS, Schiel, 2004); manual readjustments were made subsequently whenever necessary to the target word in PRAAT (Boersma and Weenink, 2012). All words that were mispronounced were excluded from the analysis. For the present study a total of 2996 words were analyzed, including 2494 /st/-clusters, 252 prototypical /s/ and 250 prototypical /ʃ/ (cf. **Table 2**).

### Experimental set-up, digitization, labeling

Spectra were extracted at the temporal midpoint between each fricative's acoustic onset and offset after applying a 256 point discrete Fourier transform with a 40 Hz frequency resolution, 5 ms Blackman window, and a frame shift of 5 ms to the target words using the Emu Speech Database system (Harrington, 2010).

The subsequent parameterization of these data involved the data reduction of each spectrum (at the sibilant's acoustic temporal midpoint in all cases) to a set of mel-scaled coefficients

using the discrete cosine transformation. More specifically, for an  $N$ -point mel-scaled spectrum,  $x(n)$ , extending in frequency from  $n = 0$  to  $N - 1$  points over the frequency range of 500–3500 Hz, the  $m$ th DCT-coefficient  $C_m$  ( $m = 0, 1, 2$ ) was calculated with the formula in (1)

$$C_m = \frac{2k_m}{N} \sum_{n=0}^{N-1} x(n) \cos\left(\frac{(2n+1)m\pi}{2N}\right) \quad (1)$$

These three coefficients  $C_m$  ( $m = 0, 1, 2$ ) encode the mean, the slope, and curvature respectively of the signal (in this case of a given sibilant's mel-scaled spectrum extracted at its temporal midpoint) to which the DCT transformation was applied (Harrington, 2010). Since  $C_0$ , which is proportional to the dB-mean across the entire spectrum, is largely irrelevant for the /s-/distinction, only  $C_1$  and  $C_2$  (the spectral slope and curvature) were used for further quantification.

We quantified the degree of neutralization of the /s-/distinction by calculating the Euclidean distances,  $E_s$  and  $E_{\text{ʃ}}$ , in the  $C_1 \times C_2$  space separately for each sibilant in the database to the Standard German speakers' /s/-centroid and to the Standard German speakers' /ʃ/-centroid, respectively. These two centroids are the positions in the  $C_1 \times C_2$  space averaged across all Standard German speakers' /s/-tokens and all Standard German speakers' /ʃ/-tokens respectively that occurred in the words from the reading condition. We then calculated for each sibilant its log-Euclidean distance ratio  $d_{\text{sib}}$ , from (2):

$$d_{\text{sib}} = \log(E_s/E_{\text{ʃ}}) = \log(E_s) - \log(E_{\text{ʃ}}) \quad (2)$$

Thus, there is one  $d_{\text{sib}}$  value per sibilant which is a relative measure: greater positive values denote a closer distance of a given sibilant to the /ʃ/-centroid; greater negative values are associated with distances closer to the /s/-centroid; and a value of zero on  $d_{\text{sib}}$  denotes that a given sibilant is equidistant in the  $C_1 \times C_2$  space between the /s/ and /ʃ/-centroids (e.g., Harrington et al., 2008; Kleber et al., 2012, for a similar methodology).

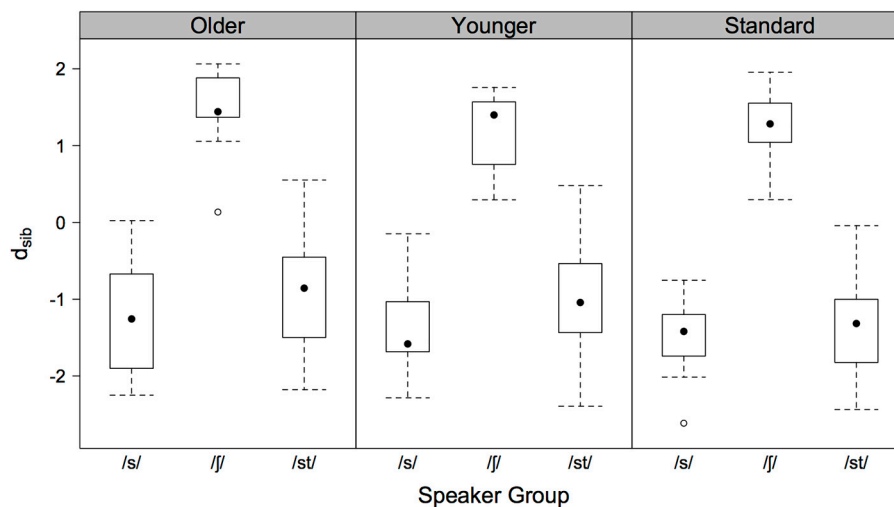
## RESULTS

**Figure 1** shows for each speaker group the log-Euclidean distance ratio,  $d_{\text{sib}}$ , for their singleton and cluster sibilants to the /s/ and /ʃ/-centroids. Negative/positive values are productions of a given sibilant closer to the /s/ and /ʃ/-centroids respectively. As **Figure 1** shows, all speaker groups produced cluster sibilants as more /s/-like, although those of older and younger AG speakers tended to be closer to the /ʃ/-centroid than those of the SG speakers: this is evident in the medians (the dots in **Figure 1**) which are higher (closer to zero) in /st/ for AG than for SG speakers.

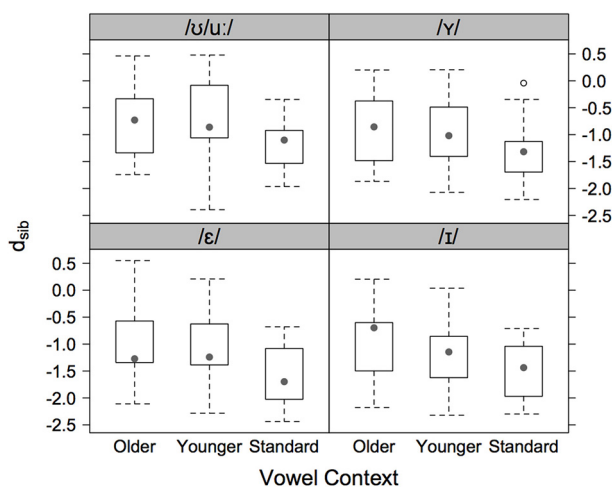
**Figure 2** shows separately for each speaker group and vowel context (/ɛ ɪ ʏ ʊ u:/)  $d_{\text{sib}}$  for the sibilants in /st/-clusters to the /s/ and /ʃ/-centroids. In these data, older AG speakers have values closest to zero: this shows that their productions were slightly more /ʃ/-like than for the other two groups. At the same time, the SG speakers always had the lowest median values such that their /st/ was closest to /s/ compared with the AG speakers. **Figure 2** also shows that the younger AG speakers' medians were between

**Table 2 | Distribution of the 2996 /s-/ʃ/-st/-sequences by age group.**

	Older AG	Younger AG	SG
/s/	66	90	96
/ʃ/	65	89	96
/st/	655	883	956



**FIGURE 1 | Boxplots of the log.** Euclidean distance ratio,  $d_{sib}$  for older AG (left), younger AG (center), and SG speakers (right). Negative values indicate productions closer to the /s/-centroid, positive values are productions closer to the /ʃ/-centroid.



**FIGURE 2 | Boxplots of the log.** Euclidean distance ratio,  $d_{sib}$  for sibilants in /st/-clusters to the /s/ and /ʃ/-centroids (older AG left, younger AG center, and SG right) and vowel context (/ɛ/ bottom left, /ɪ/ bottom right, /y/ top right, /u/ top left). Negative values indicate productions closer to the /s/-centroid, positive values are productions closer to the /ʃ/-centroid.

those of the other two groups. A mixed model with  $d_{sib}$  (the data in Figure 2) as the dependent variable and with vowel context (/ɛ ɪ y u:/) and speaker group coded for increasing order (three ordered levels: older Augsburg > younger Augsburg > Standard) and with speaker as the random factor showed a significant effect for vowel [ $\chi^2_{(1)} = 30.4, p < 0.001$ ], a significant effect for group [ $\chi^2_{(1)} = 4.7, p < 0.05$ ], and no interaction between these factors. The significant effect for group is a confirmation of the evidence in Figure 2 that there is a trend from older AG to younger AG to SG speakers for /st/ to be progressively closer to /s/.

## PERCEPTION EXPERIMENT

### METHODS

#### Participants

The participants were the same as in the production experiment. The production and perception experiments were both run in one session per speaker (always starting with the production experiment), i.e., each subject who had participated in the production experiment completed the perception experiment as well.

In order to control for the effect of biological age (i.e., for differences between groups that are not due to the dialectal background but that might come about because of an age-related diminished capacity for identifying high-frequencies that are critical for place of articulation distinctions in fricatives), we included a fourth subject group consisting of older (aged between 40 and 70 years; 7 males and 8 females) Standard German listeners. The older SG listeners were born and lived in Northern Germany (near the city of Hannover). They were tested in a quiet room at their homes. None of them reported any hearing, eye-sight, or reading problems.

#### Materials

For the perception experiment, we created two synthetic continua between /s/ and /ʃ/ using STRAIGHT Tandem (Kawahara et al., 2008). The first continuum extended between the minimal pair *Mist* (/mɪst/, *dung*) and *mischt* (/mɪʃt/, *mix*). In this context, we expected AG listeners to have difficulty perceiving the contrast, given the tendency to produce both words as homophones in this variety (we will henceforth refer to this continuum as the *ambiguous* context). The second continuum (the *unambiguous* context) extended between *vermisse* (/və'mɪsə/, first pers. sing. *miss*) and *vermische* (/və'mɪʃə/, first pers. sing. *mix*). For this continuum, we expected no difference between the groups, since the /s-/contrast is contrastively produced in both Augsburg and Standard German.

Both continua were derived from natural productions of *vermis*e and *vermisch*e spoken by a Standard German speaking phonetician. We recorded several repetitions of these two words and selected two prototypical realizations. These two selected /və'mɪsə/ and /və'mɪʃə/ sound files were morphed by adding time anchors to the /s/ and /ʃ/-sequences and setting frequency anchors for the added time anchors. This was done to get a horizontal overlap between the two sibilants. After creating a 22-step continuum between /və'mɪsə/ and /və'mɪʃə/ the mi[s/ʃ]-sequence was cut out of the created continuum. We then selected stimuli 1, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 22 (i.e., selected only 16 stimuli from the original 22 steps continuum<sup>3</sup>) for our perception experiment. After we had selected the stimuli, we prepended the synthetic mi[s/ʃ]-sequences to a following -t (to create the ambiguous continuum *Mist-mischt*) and spliced the same synthetic sequences between *ver*\_\_\_e, (to create the unambiguous continuum *vermis*e-*vermisch*e).

### Experimental procedures

The perception experiment was conducted using *Praat*'s *ExperimentMFCscript*. Listeners judged all 320 stimuli (16 stimuli × 10 repetitions × 2 contexts) in a two-alternative forced-choice identification task. The order of presenting the continua was counterbalanced, i.e., some subjects first listened to the /mɪst/—/mɪʃt/ continuum and afterwards listened to

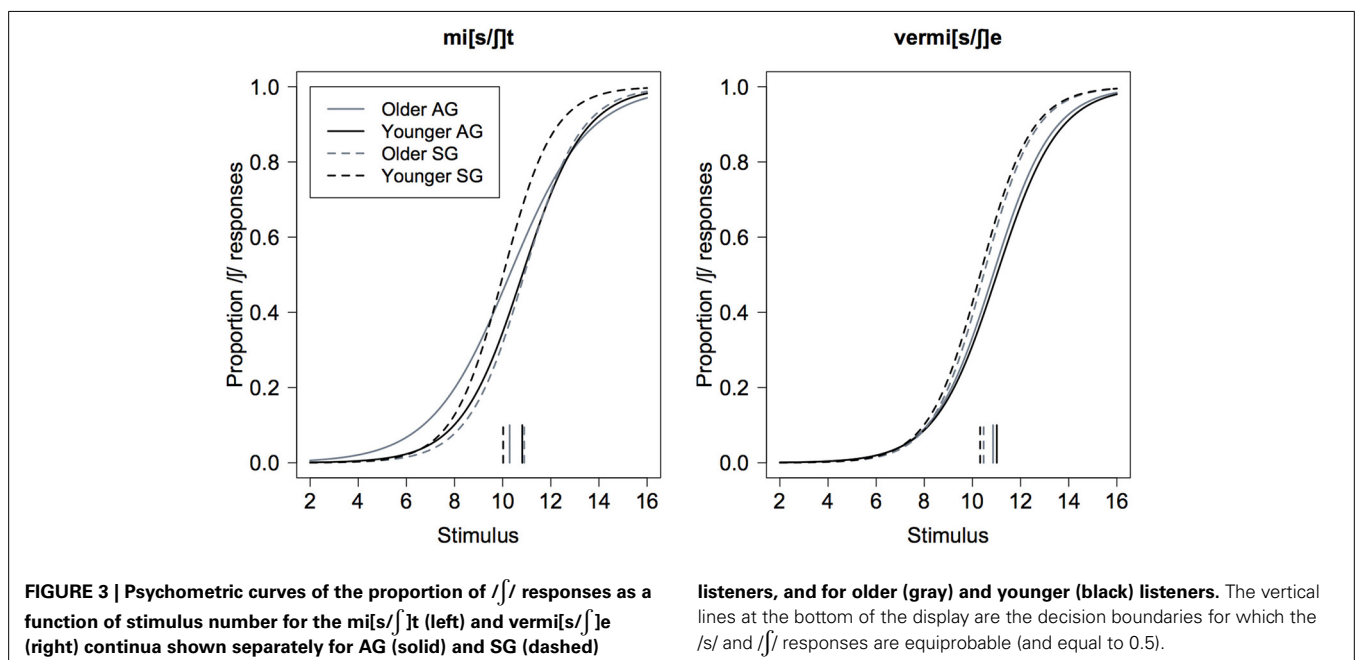
the /və'mɪsə/—/və'mɪʃə/ continuum and vice versa. All stimuli were presented to the listeners over headphones. Upon presentation of an auditory stimulus, the subject saw an orthographic representation corresponding to the minimal pair distinction. For example, the subject heard a stimulus from the /və'mɪsə/—/və'mɪʃə/ continuum and saw *vermis*e and *vermisch*e on the screen. The task then was to judge whether the stimulus sounded more like *vermis*e or *vermisch*e. The order of the stimuli was random for each participant to avoid any presentation effects. The experiment was self-paced, i.e., the next stimulus was only presented after the subject had made a decision and after a stimulus initial silence of 0.5 s. The perception experiment took about 20 min per listener.

### Data analysis

We fitted eight logistic regression models to the responses, one for each of the possible combinations of age (younger vs. older), variety (AG vs. SG), and continuum-type (mi[s/ʃ]t vs. vermi[s/ʃ]e). For each of these 8 models, the dependent variable was the binary responses (/s/ or /ʃ/), and the integer stimulus number ( $1 \leq n \leq 16$ ) was the independent (numerical) factor. The output of this analysis was used to derive psychometric curves separately by age, variety, and continuum type (Figure 3 below).

We then re-ran the same 8 logistic models, but this time included for each of them an interaction term between the stimulus and the listener: with this technique, we derived slopes, intercepts, and decision boundaries for each listener. All of the listener-specific decision boundaries fell within the range of the stimuli (i.e., between 1 and 16). However, the data from one younger AG listener on the mi[s/ʃ]t continuum and from one older Standard listener on the vermi[s/ʃ]e continuum were subsequently excluded from any further

<sup>3</sup>In order to create a continuum with a finer separation between stimuli in the middle of the continuum (i.e. the ambiguous part) we morphed a 22-step continuum. Since this fine separation was not necessary for responses to stimuli at the edges of the continuum, we discarded every second stimulus at the beginning and end of the continuum; another reason for discarding some stimuli was to shorten the duration of the perception experiment.



analyses because their slopes could not be unambiguously determined<sup>4</sup>.

## RESULTS

As the short vertical lines at the bottom of **Figure 3** show, there seems to be no systematic influence of any of the main factors on the decision boundary (the point at which the probability of /s/ or /f/ responses are equal and 0.5). On the other hand, they do have an influence on slope: in particular, the slope is clearly steeper (i.e., the psychometric functions have a more pronounced sigmoid-shape) for the Standard vs. the Augsburg listeners in both continua. In addition, the same figure suggests that there may be a steeper slope for the younger vs. older Augsburg listeners in the mi[s/f]t than in the vermi[s/f]e continuum (compare the solid black with the solid gray curves in the left panel of **Figure 3**).

The barchart of these eight slope values in **Figure 4** shows more clearly the steeper slopes for Standard vs. Augsburg listeners in all cases, as well as the steeper slope for the younger than for the older Augsburg listeners in the mi[s/f]t vs. vermi[s/f]e continuum.

**Figure 5** of the listener-specific slopes shows only some of the trends that were apparent from the analyses based on the entire population of listeners in **Figures 3, 4**. The clearest consistency is in the effect of variety: with the possible exception of older listeners on the vermi[s/f]e continuum, there is strong evidence that the slopes are steeper for the SG compared with the AG listeners. However, a comparison of the first with the second row does not

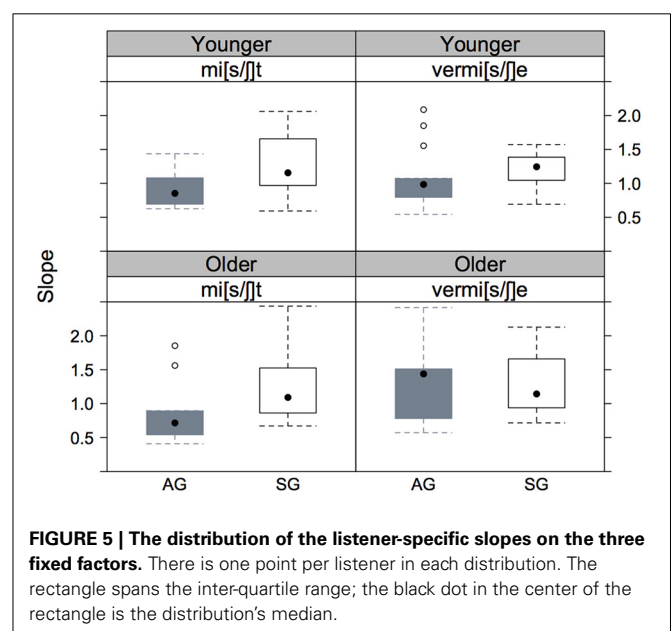
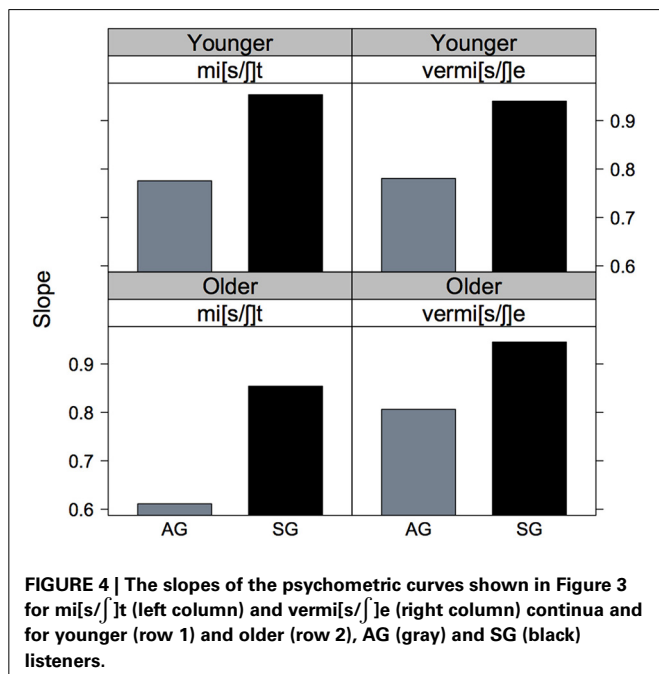
confirm the earlier observation of any influence of age group on slopes.

In order to quantify these observations further, we ran a mixed model with slope as the dependent variable, with the listener as a random factor, and with three fixed factors: age (older vs. younger), variety (AG vs. SG), and continuum-type (mi[s/f]t vs. vermi[s/f]e). The results (see also **Table 3**) showed significant main effects for variety [ $\chi^2_{(5)} = 6.3$ ,  $p < 0.05$ ] and for continuum-type [ $\chi^2 = 5.0$ ,  $p < 0.05$ ], but no effect for age group<sup>5</sup>. The results also showed no significant interactions between any of the fixed factors.

The significant effect of continuum type is to a certain extent evident in **Figure 6** in which the slopes in the vermi[s/f]e continuum have been subtracted from those in the mi[s/f]t continuum separately per listener. The null hypothesis is that the two continua do not differ on slope in which case the difference between the continua in **Figure 6** on slope should be zero. **Figure 6** shows that the median of all four distributions is above zero which means that, compatibly with the statistical analysis, the slopes were steeper on the vermi[s/f]e than on the mi[s/f]t continuum. Additionally, there was a trend for greater slope differences between continua in older Augsburg listeners (as opposed to all other speaker groups)

<sup>4</sup>This comes about for the type of data exemplified by listener KAWI in **Figure 7** for which there is only a single point (at stimulus 9) in the region of ambiguity between /s/ and /f/ responses as a result of which the slope cannot be unambiguously determined (giving rise to the error message in R “fitted probabilities numerically 0 or 1 occurred”).

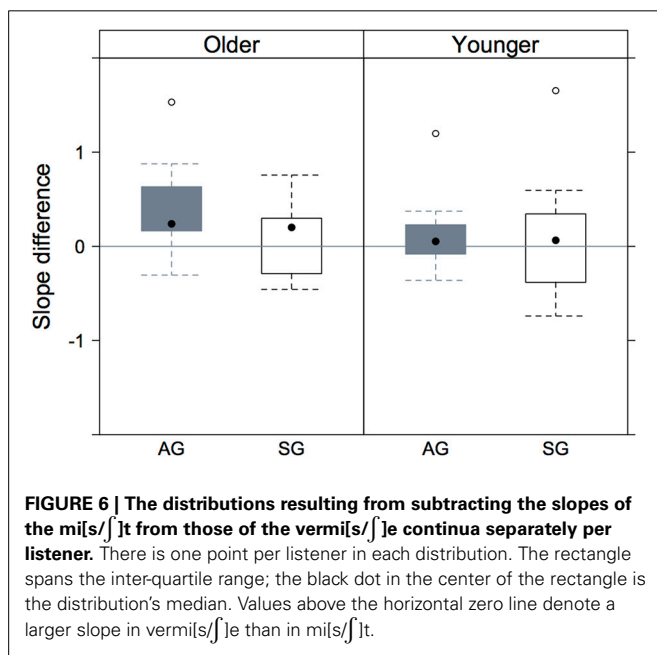
<sup>5</sup>We included a term for random intercepts for speakers which quantifies by-speaker variability in the dependent variable  $d_{sib}$ . This was because there were insufficient tokens for convergence to be obtained by additionally including random slopes. We applied a repeated measures ANOVA to the same data, in order to assess the validity of the mixed model with random intercepts only. The results showed significant influences on slope of continuum [ $F(1, 51) = 5.8$ ,  $p < 0.05$ ] and of variety [ $F(1, 51) = 6.4$ ,  $p < 0.05$ ] but not of age, and there were no significant interactions. These results are entirely consistent with those obtained from the intercept only mixed model (and also comparable in the F-statistic size and probabilities for the significant results for continuum and variety).





**Table 3 | Estimates, Standard error, and t-statistics for the independent factors in the mixed model with slope as the dependent variable.**

Factor	Estimate	St. Error	t-value
Continuum	0.41	0.15	2.82
Age	0.06	0.19	0.34
Dialect	0.42	0.19	2.22
Continuum × Age	−0.27	0.20	−1.37
Continuum × Variety	−0.35	0.20	−1.80
Age × Variety	−0.08	0.26	−0.31
Continuum × Age × Variety	0.26	0.26	0.99



since it is only for this group that the lower quartile is above zero.

## DISCUSSION

The aims of the present study were two-fold: the first was to investigate a potential sound change in progress in the Augsburg variety of German and the second to examine whether an apparently categorical sound change is gradual across generations. The motivation for this study was Bukmaier's (2010) analysis showing evidence that younger Augsburg speakers use less dialectal features such as /ft/ instead of Standard German /st/ than older Augsburg speakers. To address the present research questions, analyses of both production and perception data were mandatory. There are three main findings from this production and perception study, which are discussed below.

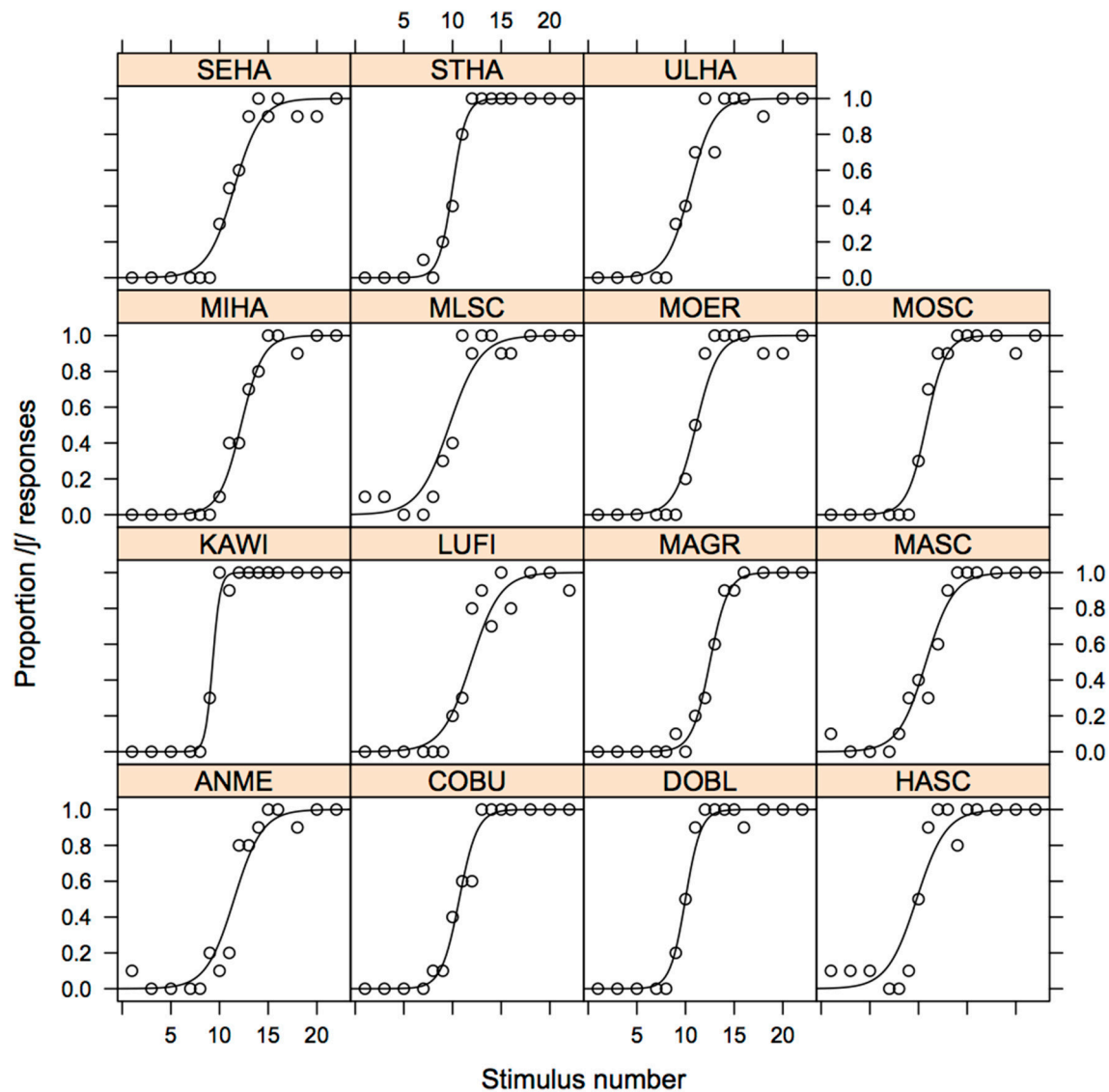
The first finding comes from the analysis of the production data showing that—although both Augsburg and Standard German speakers maintained the /s-/contrast before /t/ and produced the fricative as /s/ in this position—the sibilant in the cluster was further away from /s/ for AG compared with SG speakers. Thus, this finding supports hypothesis H1 according to

which Augsburg speakers maintain the /s-/contrast to a lesser extent in the cluster context than do Standard German speakers. As far as speaker age is concerned, hypothesis H2 predicted that the /st/-productions of younger Augsburg speakers should be between those of the older Augsburg and the Standard speakers. Our results were consistent with this hypothesis. Younger Augsburg speakers' sibilants were more /s/-like than those of their older counterparts, but not as /s/-like as those of the Standard speakers.

According to hypothesis H3, the /s-/neutralization in production in Augsburg German should have an impact on perception: that is, the /s-/contrast should not be as perceptually distinctive for AG as for SG listeners. Based on this hypothesis, we predicted that Augsburg subjects would perceive more instances of the /st-/continuum as /ft/ with the category boundary either shifted toward the /s/-end of the continuum or even with no shift from /f/ to /s/ in case of (in)complete neutralization of the contrast. Our results provide partial support for H3. On the one hand, the location of the /s-/category boundary was similar for the three groups, suggesting that there is no preference for AG listeners to perceive /f/, even though /ft/ is much more frequent in the Augsburg dialect than /st/. This result does not match Kleber et al.'s (2010) findings showing a bias in listeners' responses toward sound sequences that occur more often in a variety that the speaker is frequently exposed to. On the other hand, the results from the slopes were consistent with our hypothesis: the flatter slopes for the AG listeners are consistent with the idea that there is a greater ambiguity for AG than for SG listeners in categorizing an /s-/continuum: that is, Augsburg listeners perceived the contrast less sharply than SG listeners.

The age<sup>6</sup> effect too was less apparent in the perception than in the production data. While younger listeners' response curves appeared to be steeper and thus more categorical, this observation did not reach significance when taking differences between individuals within a speaker group into account. Therefore, hypothesis H4, which predicted that older Augsburg listeners should perceive this contrast to a lesser extent than younger Augsburg listeners is not quite supported. The results showing greater similarities across the three groups in perception than in production may be consistent with the idea that the sound change is more advanced in perception than in production. This is compatible with other findings showing a potential misalignment between the two modalities during a sound change in progress such that perception precedes production (Ohala, 1993; Kleber et al., 2012). Thus, for the present data, while older AG subjects are the most conservative of the three groups in production (because their sibilants are closest to /f/), they are similar to the younger AG listeners in how they cut up the /s-/continuum in perception. Despite the nonsignificant age-effect, our data is consistent with the view that younger speakers lead this sound change in progress from /ft/ to /st/ (Labov, 2007) since in older as opposed to younger

<sup>6</sup>Age as an artifact due to the potentially diminished perceptual capacities in older listeners can be ruled out as there were no significant differences between younger and older listeners in the standard group. That means that older speakers did not perform in general worse in sibilant perception than younger listeners.



**FIGURE 7 |** Psychometric curves of the proportion of /f/-responses for the younger Augsburg group to the stimulus of the mi[s/f]t continuum shown separately for each listener together with the raw values shown as data points.

participants' data there was (1) a trend toward flatter /st-ft/-curves (2) an apparently greater slope difference between the vermi[s/f]e and the mi[s/f]t continua, and (3) a more /f/-like pronunciation of the cluster sibilant.

In general, the Augsburg participants maintained the contrast both in perception and production in a categorical manner and thus surprisingly well. This result is probably due to Augsburg participants' awareness of the contrast between the sibilants before stops in SG. The awareness comes about because (1) they learn the standard realization in school, (2) the Augsburg variety has a phonemic /s-/contrast in intervocalic position, and (3) because they are of course exposed to the Standard German variety. Speakers also often target a standard pronunciation in a laboratory recording session. Knowledge of the contrast facilitates its production even if /ft/ is characteristic of their variety.

For example, Broersma (2005) found that Dutch listeners' performance in perceiving the final voicing contrast in English words was similar to those of English native speakers even though the voicing contrast is neutralized in final position in Dutch. She explained this finding with the listeners' capability of transferring perceptual cues from a contrast in a familiar position (such as the voicing contrast in intervocalic position in Dutch) to the same contrast in an unfamiliar position. The results from her study are consistent with our findings in perception. In addition, our production data show that speakers may also transfer these cues to the production of a contrast in an unfamiliar position—even in a blending task that is designed to obscure the aim of a study and to prevent hyperarticulation.

Studies based entirely on auditory impressions and transcriptions are not suitable for detecting the subtle differences between

speaker groups observed in the present study. It is the dialect- and age-grading found in the acoustic analyses and to a certain extent in the perception data which shows that the transfer of standard forms to the Augsburg variety is not a categorical change from /s/ to /ʃ/. Thus, the third important finding from this study is that sounds that give the auditory impression of a categorical change may nevertheless show remnants of the old or dialectal form in the acoustic signal. That is, traces of a gradual shift from the old variant toward the new variant are still present. In this respect our results are consistent with findings from physiological studies showing that articulatory traces from a segment may still be observable even though the segment is not perceptible (Poupplier and Hardcastle, 2005; Poupplier, 2007).

The findings from speech production and perception are in general consistent with previous results on gradual sound changes in German regional varieties that are most likely to evolve under the influence of the standard variety and thus can be regarded as a form of dialect leveling (Kerswill, 2003). For example, Kleber (2011) showed that, while the long/short vowel contrast tends to be neutralized before fortis stops in older Bavarian speakers, such a contrast is beginning to develop toward that of the standard variety for younger Bavarian speakers. Müller et al. (2011) report a gradual change from dialectal lenis stops toward standard fortis stops from older to younger East Franconian speakers. Similarly to East Franconian, fortis stops are lenited in Upper Saxon. Although Kleber (in press) found no age-grading in her Saxon data, she argues that the sound change in progress may be more advanced in Saxon as both older and younger Saxon speakers behaved like younger East Franconian speakers and listeners in a study by Müller et al. (2011). In addition, there was a trend toward flatter psychometric curves derived from the older Saxon listeners to a fortis-lenis continuum—similar to the data presented in this study. These forms of dialect leveling are very likely to come about because of the speakers' increasing contact with the standard language—for example, in school (Besch, 1983), via the media (cf. Stuart-Smith et al., 2013) and generally as a result of higher speaker mobility (cf. Clopper and Pisoni, 2006). The position of Augsburg in a transitional zone between Swabian and Bavarian (with Bavarian speakers patterning with Standard speakers in relation to the /s-/distinction) might further strengthen the influence of the standard variety on Augsburg German.

These forms of gradual sound changes as a consequence of dialect leveling are best explained in a usage-based model of speech perception such as the exemplar theory of speech perception (Johnson, 1997; Pierrehumbert, 2002) according to which each perceived token with all its fine phonetic detail is added to the neighborhood of the most similar exemplar in an acoustic-perceptual space of the listener's mental lexicon and where phonological categories emerge from the density distribution of the stored exemplars. The resulting exemplar cloud is not fixed to a certain point in this acoustic/auditory space, but it may shift as new exemplars that differ slightly in their acoustic make-up from the other exemplars in this cloud are added. The probability of a shift in the exemplar cloud is increased when more and more variants with properties that are auditorily at the edges of the cloud are added. Thus, in

terms of this model, a shift toward the standard variety may be caused by a progression of the cloud as more pre-consonantal /s/-exemplars from the standard variety are stored: the greater shift observed in younger AG subjects is because they are, or have been, exposed to a greater extent to SG than older subjects.

The conclusion so far is that external factors (Kerswill, 2003) cause sound change that can be associated with a general trend of dialect leveling in German regional varieties. External factors may influence sounds such that they change in a direction that would not have been predicted by phonetically motivated internal factors (Torgersen and Kerswill, 2004; Kleber, 2011; Müller et al., 2011). The present sound change in progress, however, may not only be driven by variety contact but also by internal phonetic factors. The phonetic basis of a diachronic change from /ʃ/ to /s/ before /t/ lies in the generally higher spectral peak in /ʃ/ before /t/ due to the coarticulatory effect of the alveolar stop thus pushing the sibilant toward an alveolar place of articulation. This may also explain the finding of slightly steeper vermi[s/ʃ]e than mi[s/ʃ]t-curves in all speaker groups including in the standard group (cf. Figure 6). That is, standard listeners are more variable in their choices between /s/ and /ʃ/ in an mi[s/ʃ]t-continuum because of the greater ambiguity in deciding whether a higher spectral peak is a property of the fricative itself or instead caused by the coarticulatory influence of the stop's alveolar place of articulation. Such a perceptual account would explain why a diachronic change from /s/ to /ʃ/ before /t/ is much more likely than a change from /ʃ/ to /s/.

To conclude, our findings provide evidence for a sound change in progress that affects both perception and production and which is primarily the result of the external influence of the standard variety on Augsburg German. This type of sound change patterns with a more general trend of dialect leveling in German regional varieties. Together with the results of previous studies on regional varieties of German such as East Franconian (Müller et al., 2011; Harrington et al., 2012), Saxon (Kleber, 2011, in press) and Bavarian (Kleber, 2011), these findings support the idea that the shift from one phonological category to another is gradual rather than abrupt in a context in which the categories are neutralized. In this respect, our results contribute to the longstanding debate on whether sound changes are categorical or whether phonological processes such as neutralization are complete. Phonological categories such as voiced vs. voiceless or (as in the present study) alveolar vs. post-alveolar mark endpoints of phonetic continua that span not only hyper- or hypoarticulated forms but also other forms of indeterminacy such as incomplete neutralization. Speakers produce and perceive variants along these continua. Diachronic changes may then come about when the distribution of variants along the continuum is incrementally shifted due to external factors. This idea is compatible with usage-based theories of speech perception as well as theories in which perception leads production during a sound change in progress. Future research is necessary to probe more deeply the mechanisms underlying diachronic change by investigating, for example, whether gender and social class differences or gradual shifts along phonetic continua are reinforced in certain conditions such as different prosodic contexts or speech rates,

and in certain age groups, e.g., in children during phonological acquisition.

## ACKNOWLEDGMENTS

This research was supported by an European Research Council grant number 295573 “Sound change and the acquisition of speech” (2012–2017) to Jonathan Harrington. The authors thank Ulrich Reubold for help with creating the stimuli, Lena Ehlermann and Elena Maslow for recruiting Standard German subjects and running the experiments with them, and Nikola Koczuba, Carolin Sabath, and Rebecca Stegmaier for help with segmenting and labeling the data.

## REFERENCES

- Besch, W. (1983). “Dialekt, schreibdialekt, schriftsprache, standardsprache. exemplarische skizze ihrer historischen ausprägung im deutschen,” in *Handbooks of Linguistics and Communication Science 1.2*, eds G. Ungeheuer and H. E. Wiegand (Berlin: De Gruyter), 961–990.
- Boersma, P., and Weenink, D. (2012). *Praat: Doing Phonetics by Computer [Computer Program]*. Version 5.3.08. Available online at: <http://www.praat.org/> [accessed 1 March 2012].
- Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *J. Acoust. Soc. Am.* 117, 3890–3901. doi: 10.1121/1.1906060
- Browman, C. P., and Goldstein, L. (1991). “Gestural structures: distinctiveness, phonological processes, and historical change,” in *Modularity and the Motor Theory of Speech Perception*, eds I. G. Mattingly and M. Studdert-Kennedy (Hillsdale NJ: Lawrence Erlbaum Associates), 313–338.
- Bukmaier, V. (2010). *Phonologische und morphologische Merkmale des Augsburger Dialekts*. Unpublished BA thesis, Universität Konstanz.
- Clopper, C. G., and Pisoni, D. B. (2006). Effects of region of origin and geographic mobility on perceptual dialect categorization. *Lang. Var. Change* 18, 193–221. doi: 10.1017/S0954394506006091
- Draxler, C., and Jänsch, K. (2004). “SpeechRecorder—a universal platform independent multichannel audio recording software,” in *Proceedings of 4th International Conference Language Resources and Evaluation* (Lisbon), 559–562.
- Fowler, C. A., Brown, J. M., Sabadini, L., and Weihing, J. (2003). Rapid access to speech gestures in perception: evidence from choice and simple response time tasks. *J. Mem. Lang.* 49, 396–413. doi: 10.1016/S0749-596X(03)00072-X
- Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., and Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition* 103, 386–412. doi: 10.1016/j.cognition.2006.05.010
- Harrington, J. (2010). *The Phonetic Analysis of Speech Corpora*. Oxford: Wiley-Blackwell.
- Harrington, J., Kleber, F., and Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in Standard Southern British: an acoustic and perceptual study. *J. Acoust. Soc. Am.* 123, 2825–2835. doi: 10.1121/1.2897042
- Harrington, J., Kleber, F., and Reubold, U. (2012). “The production and perception of coarticulation in two types of sound change in progress,” in *Speech Planning and Dynamics*, eds S. Fuchs, M. Weirich, D. Pape, and P. Perrier (Frankfurt: Peter Lang), 39–62.
- Harrington, J., Kleber, F., and Reubold, U. (2013). The effect of prosodic weakening on the production and perception of trans-consonantal vowel coarticulation in German. *J. Acoust. Soc. Am.* 134, 551–561. doi: 10.1121/1.4808328
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *J. Phon.* 31, 373–405. doi: 10.1016/j.wocn.2003.09.006
- Hawkins, S., and Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *J. Int. Phon. Assoc.* 35, 183–199. doi: 10.1017/S0025100305002124
- Johnson, K. (1997). “Speech perception without speaker normalization: an exemplar model,” in *Talker Variability in Speech Processing*, eds K. Johnson and J. W. Mullennix (San Diego, CA: Academic Press), 145–165.
- Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (2008). “Tandem-STRAIGHT: a temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation,” in *Proceedings ICASSP* (Las Vegas, NV), 3933–3936.
- Kerswill, P. (2003). “Dialect levelling and lexical diffusion in British English,” in *Social Dialectology. In Honour of Peter Trudgill*, eds D. Britain and J. Cheshire (Amsterdam: Benjamins), 223–243. doi: 10.1075/impact.16.16ker
- Kleber, F. (2011). *Incomplete Neutralization and Maintenance of Phonological Contrasts in Regional Varieties of Standard German*. Ph.D. dissertation, Ludwig-Maximilians-Universität München.
- Kleber, F. (in press). “Partielle neutralisierung des stimmhaftigkeitskontrastes in zwei varietäten des deutschen,” in *Thematisches Beiheft der Zeitschrift für Dialektologie und Linguistik*, eds Th. Krefeld and E. Pustka (Stuttgart: Franz Steiner Verlag).
- Kleber, F., Harrington, J., and Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Lang. Speech* 55, 383–405. doi: 10.1177/0023830911422194
- Kleber, F., John, T., and Harrington, J. (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *J. Phon.* 38, 185–196. doi: 10.1016/j.wocn.2009.10.001
- Labov, W. (1963). The social motivation of sound change. *Word* 19, 273–309.
- Labov, W. (2007). Transmission and diffusion. *Language* 83, 344–387. doi: 10.1353/lan.2007.0082
- Labov, W., Karen, M., and Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. *Lang. Var. Change* 3, 33–74. doi: 10.1017/S0954394500000442
- Lindblom, B., Guion, S., Hura, S., Moon, S., and Willerman, R. (1995). Is sound change adaptive? *Riv. Linguist.* 7, 5–37.
- Müller, V., Harrington, J., Kleber, F., and Reubold, U. (2011). “Age-dependent differences in the neutralization of the intervocalic voicing contrast: evidence from an apparent-time study on East Franconian,” in *Proceedings 12th Interspeech* (Florence), Italy.
- Niebuhr, O., and Meunier, C. (2011). The phonetic manifestation of French /s#S/ and /S#s/ sequences in different vowel contexts—on the occurrence and the domain of sibilant assimilation. *Phonetica* 68, 133–160. doi: 10.1159/000331235
- Niebuhr, O., Meunier, C., and Lancia, L. (2008). “On place assimilation in French sibilant sequences,” in *Proceedings 8th International Seminar Speech Production* (Strasbourg).
- Nübling, E. (1988). *Studien und Berichte zur Geschichts-, Mundart- und Namensforschung Bayerisch—Schwabens*. Augsburg: A. H. Konrad Verlag Weißenhorn.
- Ohala, J. J. (1981). “The listener as a source of sound change,” in *Papers from the Parasession on Language and Behavior*, eds C. S. Masek, R. A. Hendrick, and M. F. Miller (Chicago: Chicago Linguistic Society), 178–203.
- Ohala, J. J. (1993). “The phonetics of sound change,” in *Historical Linguistics: Problems and Perspectives*, ed C. Jones (London: Longman), 237–278.
- Pierrehumbert, J. (2002). “Word-specific phonetics,” in *Laboratory Phonology VII*, eds C. Gussenhoven and N. Warner (Berlin: Mouton de Gruyter), 101–140.
- Port, R., and O’Dell, M. (1985). Neutralization of syllable-final voicing in German. *J. Phon.* 13, 455–471.
- Pouplier, M. (2007). Tongue kinematics during utterances elicited with the SLIP technique. *Lang. Speech* 50, 311–341. doi: 10.1177/00238309070500030201
- Pouplier, M., and Hardcastle, W. (2005). A re-evaluation of the nature of speech errors in normal and disordered speakers. *Phonetica* 62, 227–243. doi: 10.1159/000090100
- Pouplier, M., Hoole, P. H., and Scobbie, J. (2011). Investigating the asymmetry of English sibilant assimilation: acoustic and EPG data. *J. Lab. Phonol.* 2, 1–33. doi: 10.1515/labphon.2011.001
- Renn, M., and König, W. (2009). *Kleiner Bayerischer Sprachatlas*. München: Deutscher Taschenbuch Verlag.
- Schiel, F. (2004). “MAUS goes iterative,” in *Proceedings. 4th International Conference Language Resources and Evaluation* (Lisbon), 1015–1018.
- Stuart-Smith, J., Pryce, G., Timmins, C., and Gunter, B. (2013). Television can also be a factor in language change: evidence from an urban dialect. *Language* 89, 501–536. doi: 10.1353/lan.2013.0041
- Torgersen, E., and Kerswill, P. (2004). Internal and external motivation in phonetic change: dialect levelling outcomes for an English vowel shift. *J. Sociolinguist.* 8, 24–53. doi: 10.1111/j.1467-9841.2004.00250.x
- Trudgill, P. (2004). *Dialect Contact and New-Dialect Formation: The Inevitability of Colonial Englishes*. Edinburgh: Edinburgh University Press.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 13 February 2014; accepted: 11 July 2014; published online: 31 July 2014.

Citation: Bukmaier V, Harrington J and Kleber F (2014) An analysis of post-vocalic /s-ʃ/ neutralization in Augsburg German: evidence for a gradient sound change. *Front. Psychol.* 5:828. doi: 10.3389/fpsyg.2014.00828

*This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology.*

Copyright © 2014 Bukmaier, Harrington and Kleber. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# The role of predictability and structure in word stress processing: an ERP study on Cairene Arabic and a cross-linguistic comparison

Ulrike Domahs<sup>1\*</sup>, Johannes A. Knaus<sup>2</sup>, Heba El Shanawany<sup>3</sup> and Richard Wiese<sup>4</sup>

<sup>1</sup> Institut für Deutsche Sprache und Literatur, University of Cologne, Cologne, Germany

<sup>2</sup> Department of Linguistics, Languages and Cultures, University of Calgary, Calgary, AB, Canada

<sup>3</sup> Department of German, Menoufiya University, Menoufiya, Egypt

<sup>4</sup> Institut für Germanistische Sprachwissenschaft, University of Marburg, Marburg, Germany

## Edited by:

Hubert Truckenbrodt, Centre for General Linguistics (ZAS), Germany

## Reviewed by:

Pia Knoeferle, Bielefeld University, Germany  
Marie Lallier, Basque Center on Cognition Brain and Language, Spain  
Gerrit Kentner, Goethe-Universität Frankfurt, Germany

## \*Correspondence:

Ulrike Domahs, Institut für Deutsche Sprache und Literatur, University of Cologne, Gronewaldstraße 2, 50931 Cologne, Germany  
e-mail: udomahs@uni-koeln.de

This article presents neurolinguistic data on word stress perception in Cairene Arabic, in comparison to previous results on German and Turkish. The main goal is to investigate how central properties of stress systems such as predictability of stress and metrical structure are reflected in the prosodic processing of words. Cairene Arabic is a language with a regular foot-based word stress system, leading to highly predictable placement of word stress. An ERP study on Cairene Arabic is reported, in which a stress violation paradigm is used to investigate the factors predictability of stress and foot structure. The results of the experiment show that for Cairene Arabic the internal structure of prosodic words in terms of feet determines prosodic processing. This structure effect is complemented by a frequency effect for stress patterns.

**Keywords:** metrical structure, word stress perception, Cairene Arabic, Turkish, German, P300 effect, predictability

## INTRODUCTION

Recent crosslinguistic studies on word stress perception revealed a correlation between the predictability of stress positions in a native language and the sensitivity to stress properties in second languages. In a series of studies utilizing a stress sequence recall paradigm, Dupoux, Peperkamp and colleagues found that speakers of a language with predictable word stress have difficulties to store stress information in abstract phonological representations when learning an L2 with lexical stress (e.g., Dupoux et al., 1997, 2001, 2008, 2010; Peperkamp and Dupoux, 2002; Peperkamp et al., 2010). Within a continuum of predictability ranging from predictable without exceptions to non-predictable, grades of stress-“deafness” were identified as a function of the number of exceptions from a predictable stress position. Speakers of a language with invariable stress (e.g., French) are less sensitive to stress information than speakers of a language with variable stress (e.g., Spanish). Furthermore, the more variable stress positions in a language are the more likely it is that stress information has to be lexically specified. In more recent studies, Peperkamp et al. (2010) suggest the crucial factor for stress sensitivity to be the amount of exceptional stress in a given language. The fewer cases of exceptional stress the more likely that speakers show reduced sensitivity to stress information.

So far, investigations of language specific stress representations have mainly addressed the influence of fixed vs. variable stress. The question arises what kind of stress representation has to be assumed for languages with variable stress that are said to be predictable by means of metrical structure, i.e., by predictable

parsing routines of syllables into feet. In metrical theory (e.g., Hayes, 1995) it is assumed that strong and weak syllables are grouped to either trochaic or iambic feet in which trochaic feet bear stress on the first syllable and iambic feet on the second syllable. Cairene Arabic is a trochaic and quantity-sensitive language in which bimoraic feet (consisting of either one heavy or two light syllables) are built from the left edge of a phonological word and in which the rightmost of these feet bears main stress (see Section Metrical Properties of Cairene Arabic for details, and also Hayes, 1995; Watson, 2002). Cairene Arabic is quantity-sensitive in the sense that heavy syllables build monosyllabic feet and light syllables bisyllabic ones. The position of stress varies according to the weight of the syllables and the number of feet. Thus, in contrast to languages with a fixed stress position (like final stress in Turkish; e.g., Kaisse, 1985) stress in Cairene Arabic is predictable by structure.

In order to test the effects of predictability and metrical structure, we performed a study measuring EEGs [and calculating event-related potentials (ERPs)] while native speakers of Cairene Arabic listened to correctly and incorrectly stressed words. Such a stress manipulation paradigm in an ERP study has also been applied in studies of German (Domahs et al., 2008), a language with word stress depending on metrical structure, and Turkish (Domahs et al., 2013) with mostly predictable stress. The results of both studies provide starting points to compare stress processing in a language with predictable stress (Turkish) and a language with non-predictable stress guided by metrical structure (German) with Cairene Arabic, in which stress is assumed

to be predictable as well as guided by structure. This selection of languages allows us to investigate whether the representation and processing of stress in Cairene Arabic depends mainly on the presence or absence of lexical stress specifications, on metrical structure of words or on both.

## PREVIOUS ERP STUDIES ON WORD STRESS PROCESSING

For German and Turkish word stress perception, a series of ERP experiments was performed in which participants were confronted with correctly and incorrectly stressed words of their native language (Knaus et al., 2007; Domahs et al., 2008, 2013).

The measurement of event-related potentials is suitable to investigate the online processing of certain language structures or manipulations in comparison to another condition. ERPs that are obtained via averaging processes over stimuli of the same kind and over participants are negative or positive going deflections time-locked to the stimulus onset and reflecting certain cognitive processes.

In ERP experiments on German or Turkish stress perception, trisyllabic monomorphemic words were presented auditorily, once with the correct stress pattern, and twice with the incorrect ones. The participants' task was to decide whether stress was assigned to the appropriate syllable by pressing either a "yes" or a "no" response button. The visual presentation of the target words, which immediately preceded the auditory input, helped to avoid lexical search effects, and in consequence, facilitated the decision by reducing efforts in lexical retrieval. Furthermore, the visual presentation triggered an expectation that was either met or violated in the auditory stimuli. The studies on Turkish and German demonstrated particular ERP findings, which will be summarized briefly in the following two sections.

## TURKISH

Turkish is a language with a clear default pattern: default stress is, according to many descriptions, realized on the word-final syllable (e.g., Lewis, 1967/2000; Sezer, 1981; Hayes, 1995; Kornfilt, 1997; Inkelas, 1999; Kabak and Vogel, 2001, 2011; Inkelas and Orgun, 2003; Göksel and Kerslake, 2005). The regular word-final stress pattern is quantity-insensitive, and long vowels do not attract main stress.

For the study on Turkish prosodic processing (reported in Domahs et al., 2013), a set of words with predictable final stress (e.g., *mikna'tız*; "magnet") and with exceptional lexical stress on the penultimate syllable (e.g., *ti'yatro*; "theater") was presented with either correct stress or manipulated stress on each of the other two syllables (e.g., *\*'mıknatız* or *\*mık'natız* for words with correct final stress and *\*tiyat'ro* or *\*tiya'tro* for words with correct prefinal stress). Comparisons of stress violations with correct stress conditions revealed that incorrect penultimate stress (e.g., *\*mık'natız*) evoked a positivity (between 850 and 1100 ms), while no such component occurred for the perception of items with incorrect final stress (= default stress) in words with lexical penultimate stress (e.g., *\*tiya'tro*).

Such positivity effects in evaluation tasks have been suggested to reflect sensitivity to a deviant structure with an amplitude being correlated with the degree of abnormality (e.g., Picton, 1992; Coulson et al., 1998): The less likely a metrical structure

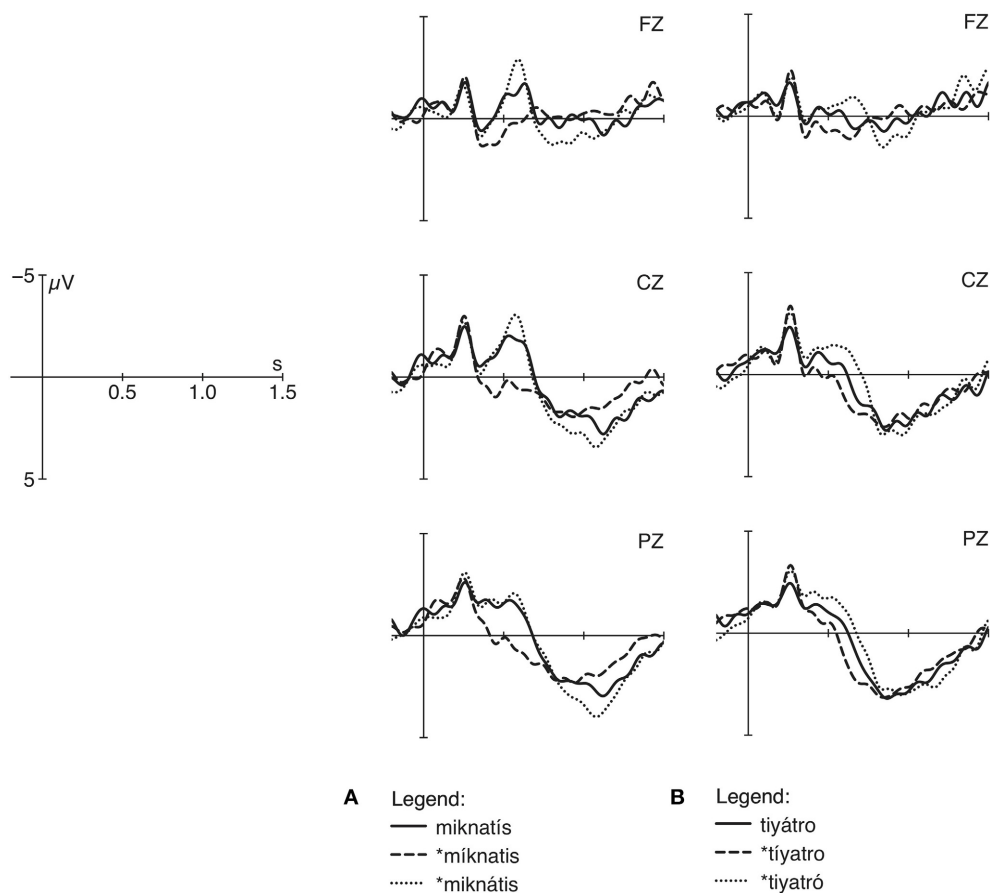
the more pronounced the positivity effect. In the literature, this task-related component has been labeled P300 (e.g., Picton, 1992; Coulson et al., 1998), P600 (e.g., Marie et al., 2011; Schmidt-Kassow et al., 2011a,b) or LPC (e.g., Rugg and Nagy, 1989). The P300 reflects decision-making processes where the reduction of the amplitude indicate that stimulus information is not clear enough. Thus, this component reflects indirectly the grammaticality in stimulus categorization (e.g., Nieuwenhuis et al., 2005).

The different ERP results for deviating stress patterns in Turkish is depicted in **Figure 1**. In words with correct final stress (**Figure 1A**), both violations produce a late positivity if compared with the correct condition. The latency of the positivity, however, differs due to the fact that the position of stressed syllables, which are decisive for the identification of stress patterns, varies. In contrast to **Figures 1A,B** depicts a positivity effect for violations with initial stress in words with canonical penultimate stress, but no positivity effect for violations involving final stress. The asymmetrical patterning of positivity effects for the two word sets suggests that Turkish participants are sensitive to lexical stress patterns but insensitive to default stress, because violations with lexical stress patterns are perceived as less likely in contrast to violations with the default stress. Thus, our findings support and complement findings by Peperkamp et al. (2010) for languages with predictable stress.

In addition to the P300 effect, an N400 effect, a negative going deflection around 250 and 500 ms post-stimulus onset, was obtained for violations with final stress. This effect was interpreted to reflect brain responses to an unexpected stimulus that produce higher costs in lexical retrieval (for a review of the N400 component see Kutas and Federmeier, 2011). Note that a shift from the lexical non-final stress position (*ti'yatro*) to the default (*\*tiya'tro*) involves a violation of a lexical stress specification. It is most remarkable that the Turkish participants showed this negative deflection mirroring the violation of an expected stress pattern while they had difficulties to classify the incorrect default stress as violating. The difficulties were not only indicated by a lack of a P300 effect but also by high error rates in the behavioral data.

## GERMAN

German monomorphemic words allow for final, penultimate, or antepenultimate stress. Which pattern to occur cannot be adequately predicted by means of stress rules. Though the stress position itself is considered not predictable, the underlying prosodic structure can be determined mostly on the basis of the weight of the final syllable. In most accounts of German phonological words, trochees are built in a right-to-left manner (Eisenberg, 1991; Wiese, 1996; Féry, 1998; Janssen, 2003). In words with a heavy final syllable (*Vitamin*—((vi.ta)<sub>F</sub>(mi:n)<sub>F</sub>)<sub>ω</sub>), the final syllable constitutes a non-branching foot (a moraic trochee), and in words with a light final syllable, the final syllable constitutes the weak syllable of a bisyllabic trochee. Thus, trisyllabic words varying in the structure of the final syllable consist of either two feet ((σσ)<sub>F</sub>(σ)<sub>F</sub>)<sub>ω</sub> or one foot (σ(σσ)<sub>F</sub>)<sub>ω</sub> (for such an analysis see Janssen, 2003; Domahs et al., 2008, 2014; Knaus and Domahs, 2009; Röttger et al., 2012).

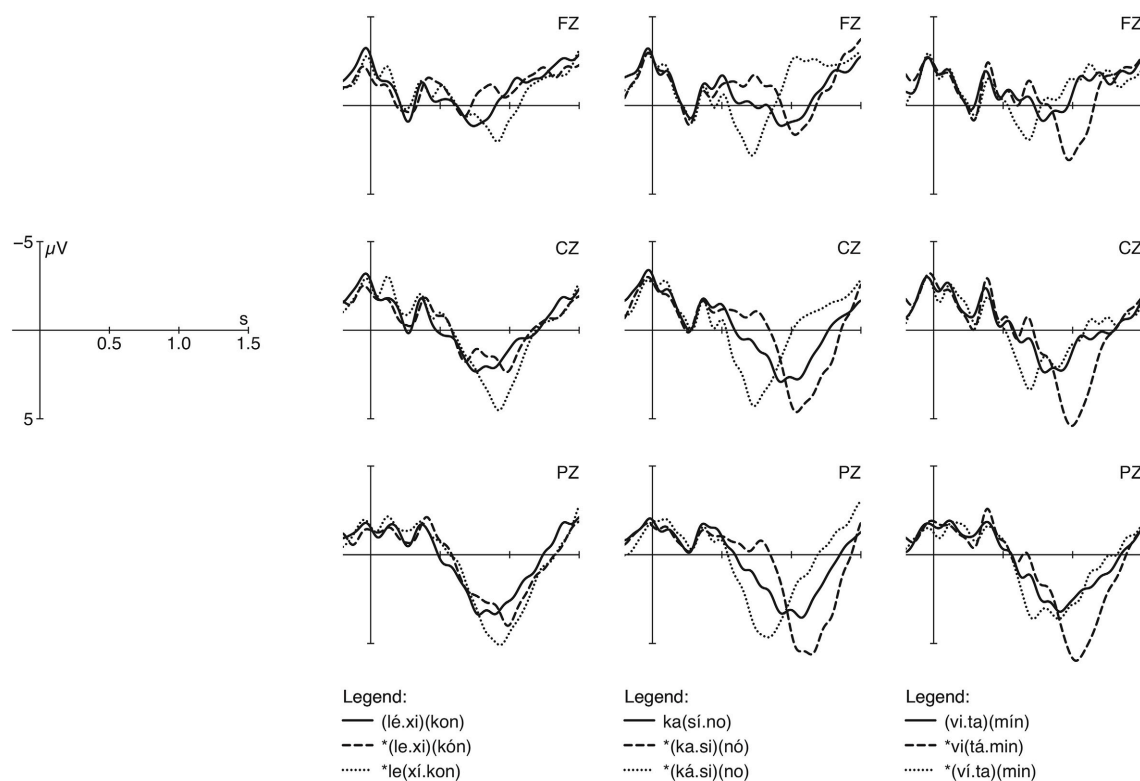


**FIGURE 1 | Grand-average curves of correctly and incorrectly stressed Turkish words (see also Domahs et al., 2013) measured at midline electrodes.** Correct words are illustrated by solid lines, incorrect antepenultimate stress by dashed lines and incorrect penultimate/final stress by dotted lines.

The experiment on German word stress evaluation (reported in Domahs et al., 2008) revealed different ERP patterns compared to the findings on Turkish. Again words with antepenultimate, penultimate and final stress were recorded with correct and deviating stress on each of the other two syllables. In contrast to Turkish, no effect of default stress was found, although several proposals assume penultimate stress to be the default stress pattern (as in *Ka'sino*; “casino”) in German. If the penultimate stress were the default stress pattern, we would expect this pattern not to evoke a late positivity when used incorrectly. However, incorrect penultimate stress in trisyllabic words with either correct final (e.g., *\*Vi'tamin* instead of *Vita'min*; “vitamin”) or initial stress (e.g., *\*Le'xikon* instead of *'Lexikon*; “lexicon”) evoked enhanced positivity effects (between 900 and 1150 ms) showing that participants can decide clearly that this stress is incorrect (see Figure 2). However, comparisons between correct and incorrect conditions revealed another form of asymmetric results regarding the occurrence or non-occurrence of a P300 component in German stress perception: stress violations produce enhanced positivity effects whenever the stress derivation leads to a change in foot structure (e.g., *\*vi('ta.min)*<sub>F</sub> instead of *(vi.ta)('min)*<sub>F</sub>), but not if the foot structure is maintained (e.g., *\*('vi.ta)(min)*<sub>F</sub> instead of *(vi.ta)('min)*<sub>F</sub>). In contrast to Turkish, it is not the

main stress position but rather the internal prosodic structure of words that is more or less predictable and has an impact on the processing of word stress (see Janssen, 2003; Domahs et al., 2014). In addition, behavioral data (error rates) indicate that German participants are sensitive to stress manipulations and identify incorrect stress with high accuracy, while Turkish participants recognized violations involving default stress at chance level only.

In the present paper, we examine a third type of language, Cairene Arabic, with a predictable and foot based stress system. Strictly bimoraic feet are built from left to right and the rightmost foot receives main stress (see below Section Metrical Properties of Cairene Arabic). Hence, Cairene Arabic is situated between the Turkish and German system by having predictable word stress like Turkish, but varying positions of word stress due to quantity sensitive foot formation like German. The main goal was to see whether speakers of Cairene Arabic are insensitive to the very predictable stress positions in their language (as Turkish participants have been shown to be insensitive to the predictable stress pattern), or whether asymmetrical ERP results occur along the lines of metrical structure (stress derivation that change the structure produce P300 effects and those that maintain structure not). To test this, trisyllabic words with penultimate and final stress were



**FIGURE 2 |** Grand-average curves of correctly and incorrectly stressed German words (see also Domahs et al., 2008) measured at midline electrodes.

compared in two conditions each: (i) penultimate words with one foot [e.g., *va('nil)Fja*; “vanilla”; in the following word type 1] and with two feet [e.g., *(mus)F('taf)Ffa*; “hospital”; in the following word type 2] and (ii) finally stressed words with a bisyllabic initial foot and a monosyllabic final foot [e.g., *(vi.ta)F('mi:n)F*; in the following word type 3] and with two monosyllabic feet [e.g., *ki(ris)F('ta:l)F* “crystal”; in the following word type 4]. If structure licenses stress positions, we should find that deviating stress realized on a strong syllable of a foot produces less pronounced positivities compared to deviating stress on a weak syllable (for instance, incorrect antepenultimate stress in words of the structure *(mus)F('taf)Ffa* should evoke less pronounced effects compared to incorrect antepenultimate stress in words of the structure *va('nil)Fja*).

Before we continue to present the experiment on Cairene Arabic we would like to introduce the main characteristics of the Cairene Arabic stress system.

### METRICAL PROPERTIES OF CAIRENE ARABIC

The Cairene Arabic dialect of Arabic is the most widely spoken language in Egypt. Half of the population speaks the Cairene Arabic dialect as its first language. Note that Cairene Arabic is a spoken language (though also written forms exist), while the literary language of Egypt is Standard Arabic (Woidich, 2006).

Cairene Arabic is not only the most widely spoken dialect in Egypt, it is also the best described Arabic dialect, particularly as regards its metrical structure. In the literature,

pre-generative (Harrell, 1960; Mitchell, 1960), generative (Hayes, 1995; Watson, 2002), and typological accounts (Hulst van der and Hellmuth, 2010) exist, which all identify Cairene Arabic as a quantity-sensitive language in which the parsing of syllables into feet is sensitive to syllabic weight: a super-heavy final syllable with long vowels followed by a consonant (CVVC) receives main stress, otherwise a heavy penult with a long vowel or a short vowel followed by a consonant is stressed or a light antepenult in words ending in three light syllables (open syllables with short vowels). According to McCarthy (1979), bimoraic trochees consisting of either one heavy syllable or two light syllables are built in a left to right manner. In (1) examples for words with final, penultimate, and antepenultimate stress are given.

- (1)
  - (a) **final stress**  
[ga'to:] “cake”, [vita'mi:n] “vitamin”, [kiris'ta:l] “crystal”
  - (b) **penultimate stress**  
[be:tak] “your house”, [va'nilja] “vanilla”, [mus'taf fa] “hospital”
  - (c) **antepenultimate stress**  
[k'azino] “casino”, [san'timitir] “centimeter”

The syllable in Cairene Arabic consists obligatorily of a single onset consonant followed by a short or long vowel. The coda maximally includes two consonants, but only one consonant in word-medial position. Syllable weight is important for the foot

formation in Cairene Arabic because feet consist of minimally and maximally two moras, a unit proposed to define syllable weight (e.g., Hyman, 1985). Accordingly, syllables with a long vowel or a short vowel followed by a coda consonant (two moras) are heavy and syllables with a short vowel (one mora) are light. For word final syllables different conditions must be met for a syllable to be heavy because the final consonant is analyzed to be extrametrical, i.e., does not contribute to syllable weight. Therefore, a final syllable is heavy if it consists of a long vowel or a short vowel followed by two consonants. These properties of heavy and light syllables guide the foot formation of phonological words in Cairene Arabic. In (2), the analysis according to Hayes (1995: 69/70; following McCarthy, 1979) is summarized.

- (2) Rules for Cairene Arabic stress
  - (a) word final consonants are extrametrical: C → <C> / \_\_\_\_]word
  - (b) foot construction: Build up bimoraic trochaic feet from left to right  
No degenerate feet!
  - (c) word layer construction: Group feet into a right-headed word constituent  
(End Rule Right)

We also note that there are other types of evidence for the bimoraic trochee in this language although secondary stress corresponding to a foot not carrying word stress has been reported to be absent (Watson, 2002, ch. 5): the word in Cairene Arabic minimally consists of a bimoraic foot. Furthermore, there is a productive pattern for nick names or hypocoristics in which names of any prosodic shape are truncated to a bimoraic foot, see examples in (3).

(3) full form	hypocoristic
Fahd	<sup>1</sup> do.do
Karim	<sup>1</sup> Ki.ki
Shaimaa	<sup>1</sup> ʃo.ʃo
Mostafa	<sup>1</sup> S'a.s'a

The present study is designed to investigate whether the foot structure as proposed in metrical analyses of Cairene Arabic are psychologically real and used during the processing of lexical words.

## ERP EXPERIMENT ON CAIRENE ARABIC

The method used in the present ERP-experiment was adopted from the ones on German and Turkish reported in Domahs et al. (2008) and Domahs et al. (2013). Similar to the previous studies, participants were confronted with correctly and incorrectly stressed words and instructed to judge the correctness of the stress patterns. Given the results on German, this stress-violation paradigm utilizing explicit judgments of stress proved to be suitable to investigate factors involved in prosodic processing of words. In particular, this method enables to identify potential stress positions irrespective of the correct one. In the following, we will present the experiment on Cairene Arabic in more detail

and compare the results with those obtained from German and Turkish participants.

**CAIRENE ARABIC**

The aim of the present experiment is to test whether (i) native speakers of Cairene Arabic are sensitive to stress manipulations and (ii) whether the processing of stress manipulations is influenced by foot structure. For this purpose, participants were presented with correctly and incorrectly stressed trisyllabic words differing in syllable and foot structure.

### Participants

Twenty-three right-handed native speakers of Cairene Arabic (20 men) were recruited for participation at the University Marburg, all of which having normal or corrected-to-normal vision and no hearing deficits. The participants' age ranged from 26 to 45 (mean age 32). All participants were born and raised monolingually in and around Cairo in Egypt, all from the Cairene Arabic dialect region. The participants' language skills comprised of second language knowledge of English, German, French, or Spanish. All participants stated to have been raised monolingually with Cairene Arabic as ambient language, and had been in Germany for 36 month in mean before participation, ranging from 1 month up to 7 years. Participants were instructed in Cairene Arabic to ensure that participants are well informed. Each participant was paid for his/her contribution. The data sets of three participants had to be excluded due to missing responses, left-handedness or excessive movement artifacts.

Note that a balanced proportion of women and men could not be obtained due to the fact that participation would have required removing the headscarf.

### Material

In order to be able to investigate whether the foot structure constrains the processing of stress shifts, we investigated four word types that differ in foot structure, as summarized in **Table 1**. Words with structure 1 and 2 are canonically stressed on the penultimate syllable and consist of heavy penultimate syllables with either long or short vowels followed by a consonant (for the sake of clarity only rhyme structures are illustrated, i.e., a structure CVC is mentioned as VC) and the first syllable is either footed or not, words with structure 3 and 4 are canonically stressed on the final syllable and contain super heavy final syllables. In structure 3, the first two syllables constitute a bisyllabic foot while in structure 4 the heavy penult constitutes a monosyllabic foot.

In words with canonical penultimate stress (structure 1 and 2), the question is whether stress moved from penultimate syllable to antepenultimate syllable produce less pronounced P300 effects when the antepenultimate syllable is head of a foot (structure 2) in comparison to unfooted (structure 1). In words with canonical final stress (structure 3 and 4), either the antepenultimate syllable (structure 3) or the penultimate syllable (structure 4) is the head of a foot and therefore a potential landing site for stress. Though the existence of secondary stress is disputed in Cairene Arabic, the question arises whether words are exhaustively parsed into feet and whether heads of feet are stressable in contrast to weak syllables of feet.



For each type of trisyllabic words, a set of 15 monomorphemic items (as given in Appendix) was selected and recorded by a female native speaker of Cairene Arabic in a sound-proof booth (44 kHz, 16 bit, mono). Each word was realized in the

**Table 1 | Conditions and material.**

Structure	Conditions	Examples
1 V(VC)V	Correct PU stress	va.'nil.ja
	Incorrect APU stress	*'va.nil.ja
	Incorrect final stress	*va.nil.'ja
2 (VC)(VC)V	Correct PU stress	mus.'taʃ.fɑ
	Incorrect APU stress	*'mus.taʃ.fɑ
	Incorrect final stress	*mus.taʃ.'fɑ
3 (V.V)(VVC)	Correct final stress	vi.ta.'mi:n
	Incorrect APU stress	*'vi.ta.mi:n
	Incorrect PU stress	*vi.'ta.mi:n
4 V(VC)(VVC)	Correct finals stress	ki.ris.'ta:l
	Incorrect APU stress	*'ki.ris.ta:l
	Incorrect PU stress	*ki.'ris.ta:l

correct and in the two incorrect conditions (see **Table 1**). In order to ensure that incorrect stresses were not produced in an exaggerated manner, correct and incorrect words with the same stress pattern were recorded in a randomized list. The phonetic parameters of duration, intensity, and F0 of each stress pattern were compared between correct and incorrect conditions (e.g., between correct *kiris'ta:l* and incorrect \**vanil'ja*, see **Table 2** with mean values for each stress patterns) showing that incorrect and correct stress realizations of a certain stress pattern differ significantly only with respect to duration because correct and incorrect conditions differ in syllable structure (e.g., *kiris'ta:l* ends in a super heavy syllable while \**vanil'ja* does not; for the statistical analyses of phonetic parameters see **Table 2**). But crucially, correct and incorrect versions of each stress pattern do not differ regarding F0 and intensity.

Furthermore, the stimuli were not spoken in isolation but embedded in the following carrier sentence:

- (3) howa lazem ye?ool **vitami:n** delwa?ti “He has to say Vitamin now!”

**Table 2 | Mean values (SD in parentheses) of phonetic parameters fundamental frequency (F0 in Hz), duration (ms), and intensity (dB) as well as repeated measures ANOVAs on the factor CORRECTNESS (correct vs. incorrect) per stress pattern.**

Stress pattern	Condition	Parameter	1st syllable	2nd syllable	3rd syllable
Antepenultimate stress	Correct (filler items)	F <sub>0</sub>	238 (8.7)	204 (6.0)	168 (29.9)
		Duration	226 (55)	196 (33)	278 (55)
		Intensity	53.3 (3.3)	51.2 (4)	39.4 (3.7)
	Incorrect	F <sub>0</sub>	238 (11.0)	199 (9.3)	180 (25.6)
		Duration	279 (80)	233 (51)	300 (64)
		Intensity	53.3 (5)	47.4 (5.5)	39.5 (4.2)
Penultimate stress	Correct	F <sub>0</sub>	216 (5.6)	228 (9.6)	189 (15.5)
		Duration	265 (68)	358 (67)	339 (52)
		Intensity	49.1 (4.9)	55.7 (3.6)	41.6 (3.2)
	Incorrect	F <sub>0</sub>	219 (11.3)	231 (7.5)	183 (25.9)
		Duration	216 (62)	425 (62)	386 (81)
		Intensity	51.2 (6.5)	51.8 (4.7)	40.1 (3.9)
Final stress	Correct	F <sub>0</sub>	221 (8.0)	218 (5.2)	214 (7.2)
		Duration	172 (57)	255 (68)	629 (73)
		Intensity	52 (4.5)	52 (5.9)	48.6 (3.5)
	Incorrect	F <sub>0</sub>	218 (9)	216 (5)	215 (5.9)
		Duration	224 (78)	232 (63)	539 (77)
		Intensity	50.7 (5.3)	49.7 (5.7)	48.4 (3.4)

#### REPEATED MEASURES ANOVA

Antepenultimate stress	F <sub>0</sub>	$F_{(1, 19)} = 2.49; p > 0.13$
	Duration	$F_{(1, 19)} = 15.4; p < 0.001^{***}$
	Intensity	$F_{(1, 19)} = 3.2; p > 0.08$
Penultimate stress	F <sub>0</sub>	$F_{(1, 29)} < 1$
	Duration	$F_{(1, 29)} = 11.0; p < 0.003^{**}$
	Intensity	$F_{(1, 29)} < 1$
Final stress	F <sub>0</sub>	$F_{(1, 29)} = 2.3; p > 0.14$
	Duration	$F_{(1, 29)} = 8.4; p < 0.008^{**}$
	Intensity	$F_{(1, 29)} < 1$

Significant results are indicated by <\*>

Significant results at 1% level are indicated by <\*\*\*> and at 0.1-level by <\*\*\*>.

The carrier sentence was identical for each critical stimulus and included the stimulus in a citation-like context bearing nuclear stress. The carrier sentence avoids a list reading and a pitch fall at the end of the critical words.

Each of the 15 items per word condition was presented in the correct and in the two incorrect conditions. To increase the number of items per condition, each version of a stimulus was presented twice. Thus, the total number of critical items was 4 (word types)  $\times$  15 (individual items)  $\times$  3 (stress patterns)  $\times$  2 (repetitions) resulting in 360 tokens. In addition, 80 trials including words with correct antepenultimate stress were included as filler. This was done to ensure that each stress pattern occurred in correct and incorrect conditions, and that the number of correctly and incorrectly stressed words was balanced.

### Procedure

Participants were seated in front of a computer screen in a sound-proof room. In each trial they were confronted with the visual presentation of an experimental item followed by the auditory presentation of the same item. The participants' task was to decide as accurately as possible whether the auditory stimuli were correctly stressed or not by pressing a response key of a push-button box. The task required the participants to activate internal stress representations (from the written input) and to compare these representations with stress information in the auditory presentation.

Each trial started with a fixation cross that appeared for 500 ms. An experimental item was then presented visually for 900 ms, followed by a blank screen for 250 ms before the auditory presentation of the stimulus started. The mean duration of the sentences was 3.9 seconds. Throughout the auditory presentation, the participants were asked to fixate on a cross in the center of the screen to avoid eye movement artifacts while listening. After the offset of each sentence, a question mark appeared on the screen and remained there until a yes or no button was pressed with a timeout of 2000 ms. Responses were given after the appearance of the question mark, but not immediately while listening to the critical items, to avoid movement artifacts. The assignment of thumbs to the yes and no buttons was counterbalanced across participants. During the answering period and the following intertrial interval of 3000 ms, the participants were allowed to blink and to rest their eyes. The experiment was controlled by the Presentation software (Version 15; Neurobehavioral Systems).

The stimuli appeared in eight experimental blocks consisting of 55 stimuli each, preceded by a short practice phase. Experimental and filler items were presented in pseudo-randomized order, each word appearing only once within each block. The order of blocks was varied for each participant to avoid sequence effects. The entire duration of the experimental session was approximately 60 min.

### Data acquisition and analyses

#### (a) Behavioral Data

During each trial accuracy and reaction time data were measured. For statistical analyses, only the accuracies of judgments were calculated because response latencies were measured after the offset of the sentences with a delay of approximately 880 ms.

The accuracy scores were calculated for each participant and condition and for each stimulus and condition.

In two repeated measures ANOVAs, the factors FOOT STRUCTURE (two different structures) and STRESS POSITION (antepenultimate, penultimate, and final) were analyzed in a  $2 \times 3$  design for words with canonical penultimate and canonical final stress separately. We calculated two separate ANOVAs due to the fact that the structure conditions for words with either penultimate or final stress vary systematically.

#### (b) ERP Data

An electroencephalogram (EEG) was recorded from overall 24 Ag/AgCl electrodes via a *BrainVision* (Brain Products) amplifier. Four electrodes measured the electro-oculogram (EOG), i.e., horizontal and vertical eye movements. The reference electrode was placed at the left mastoid. EEGs were re-referenced off-line to both mastoids. The C2 electrode served as ground. The head electrodes were mounted on an elastic cap (Easy Cap). EEG and EOG were recorded with a sampling rate of 500 Hz and filtered offline with a 0.3 to 20 Hz bandpass filter. All electrode impedances were kept below 5 k $\Omega$ . Prior to data analysis, all individual EEG recordings were automatically and manually scanned for artifacts from eye or body movements and muscle artifacts. In total, 7.5% of the data with an amplitude change of more than 40  $\mu$ V had to be excluded from analysis.

Averages were calculated per participant and condition starting from the onset of the auditory stimulus up to 1500 ms. For words with correct penultimate or final stress, incorrect conditions were compared with correct conditions. In analogy to earlier studies (Domahs et al., 2008, 2013), time-windows were chosen by visual inspection for the two sets of words with canonical penultimate and final stress pattern separately because the latency of effects reflecting the evaluation of stress patterns and decision making seem to depend on the position of the stressed syllable. Therefore, effects measured for words with incorrect antepenultimate stress occur earlier than effects found for incorrect penultimate and final stress.

Furthermore, violations with penultimate and final stress evoked a biphasic pattern consisting of a negativity followed by a positivity, while violations with antepenultimate stress evoked only a positivity. This lack of a negativity is due to the fact that the positivity occurs within the negativity time-window. **Table 3** provides an overview of time-windows per word type and incorrect stress condition. For each time window, a general analysis of variance with repeated measures (ANOVA) was calculated for words with canonical penultimate and canonical final stress separately over the factors FOOT STRUCTURE (the two different foot structures per correct stress pattern; structure 1 and 2 are compared for words with canonical penultimate stress and structure 3 and 4 for words with canonical final stress) correctness (correct vs. incorrect) and region (frontal, central, parietal). Region is defined as a three-level factor with the values frontal (including F3, Fz, F4), central (including C3, Cz, C4), and parietal (including P3, Pz, P4).

### Results

#### (a) Behavioral Data

In the analyses of accuracy scores, the aim was to investigate whether specific conditions were more error-prone than others.

**Table 3 | Time-windows for statistical analyses.**

Stress condition	Structure	Violation type	Time-windows	
			Negativity effect	Positivity effect
Correct penultimate stress	1	Antepenultimate stress	–	350–600 ms
	V('VC)V	Final stress	400–550 ms	800–1150 ms
	2	Antepenultimate stress	–	350–600 ms
	(VC)('VC)V	Final stress	400–550 ms	800–1150 ms
Correct final stress	3	Antepenultimate stress	–	300–650 ms
	(V.V)('VVC)	Penultimate stress	400–480 ms	550–850 ms
	4	Antepenultimate stress	–	300–650 ms
	V(VC)('VVC)	Penultimate stress	400–480 ms	550–850 ms

For each word type, the correct conditions with two different foot structures are compared to each incorrect condition. Time windows are given for negativity and positivity effects.

A repeated measures ANOVA of arcus-sinus transformed accuracy scores was calculated over the factors FOOT STRUCTURE (two different structures) and STRESS POSITION (antepenult, penult, and final stress) for the two sets of words with either canonical penultimate or final stress, and pairwise *t*-tests comparing correct with incorrect stress and both incorrect conditions per word set. **Figure 3** depicts the mean accuracy scores for all conditions.

Generally, speakers of Cairene Arabic are accurate with their judgments for more than 80% in each condition. This finding suggests that they are in principle sensitive to the presented stress manipulations. However, the accuracy for all conditions differs slightly, and as is illustrated in **Figure 3**, the mean accuracy for conditions with incorrect antepenultimate stress is lower compared to other conditions. Repeated measures ANOVAs and paired *t*-tests are calculated for words with canonical penultimate and final stress separately (see **Table 4**).

Analyses for words with correct penultimate stress yield a main effect for the factors FOOT STRUCTURE and STRESS POSITION as well as an interaction of both factors. *Post-hoc t*-tests comparing mean accuracies of the correct condition with each incorrect condition and of both incorrect conditions revealed a significant difference between two incorrect conditions. This holds for both word types with canonical penultimate stress.

Analyses for words with correct final stress yield a main effect for the factor STRESS POSITION and an interaction of the factors FOOT STRUCTURE and STRESS POSITION. *Post-hoc t*-tests revealed a significant difference between mean accuracy for incorrect antepenultimate stress and incorrect penultimate stress in words of the structure (V.V)(V:C) but not in words of the structure V(VC)(V:C). Overall, the analyses suggest that conditions with incorrect antepenultimate stress are more error-prone than correct conditions and other incorrect stress conditions. This could be interpreted as an uncertainty toward words containing incorrect antepenultimate stress. Note that accuracies for correct words with antepenultimate stressed (filler condition) scored high with 98% correct responses.

#### (b) ERP Data

For the analyses of mean voltage changes induced by stress manipulations, we calculated for each set of words with either

canonical penultimate or final stress whether each of the two incorrect conditions differ significantly from the correct condition and whether the foot structure influences the processing of incorrectly stressed words. **Figure 4** shows the grand averages at midline electrodes for the four word types. Generally, we observed positivity effects for stress deviations involving antepenultimate stress and a biphasic ERP pattern for violations with penultimate or final stress. As noted in Section Data Acquisition and Analyses, effects for violations with antepenultimate stress occur in earlier time-windows compared to effects for violations with penultimate or final stress. Therefore, mean voltage changes for the processing of separate stress deviations were analyzed in different time windows. Appendix provides an overview of statistical analyses. In the following, the results are presented for each set of words with either penultimate or final stress separately.

**Words with canonical penultimate stress.** Violations with antepenultimate stress (dashed line in **Figures 4A,B**) produced a positivity effect between 350 and 600 ms in the two word types with canonical penultimate stress. A main effect for the factors CORRECTNESS and REGION and an interaction for FOOT STRUCTURE  $\times$  CORRECTNESS  $\times$  REGION occurred. *Post-hoc* analyses confirm significant differences between correct and incorrect antepenultimate stress in each region and for each structure (see **Table A2A**).

Violations with final stress in words with canonical penultimate stress (dotted line in **Figures 4A,B**) evoked a biphasic ERP pattern consisting of a negativity effect between 400 and 550 ms and a positivity effect between 800 and 1150 ms. For the negativity, repeated measures ANOVAs revealed a main effect for the factors CORRECTNESS and REGION and an interaction for REGION  $\times$  FOOT STRUCTURE for which *post-hoc* analyses exhibited no significant structure effects in the three regions (see **Table A2B**). For the positivity effect, a main effect for the factors CORRECTNESS and REGION and a three-way interaction was obtained. *Post-hoc* analyses show that mean voltages differ significantly between correct and final stress in parietal region for words of the structure V(VC)V, and in centro-parietal region for words of the structure (VC)(VC)V (see **Table A2C**).

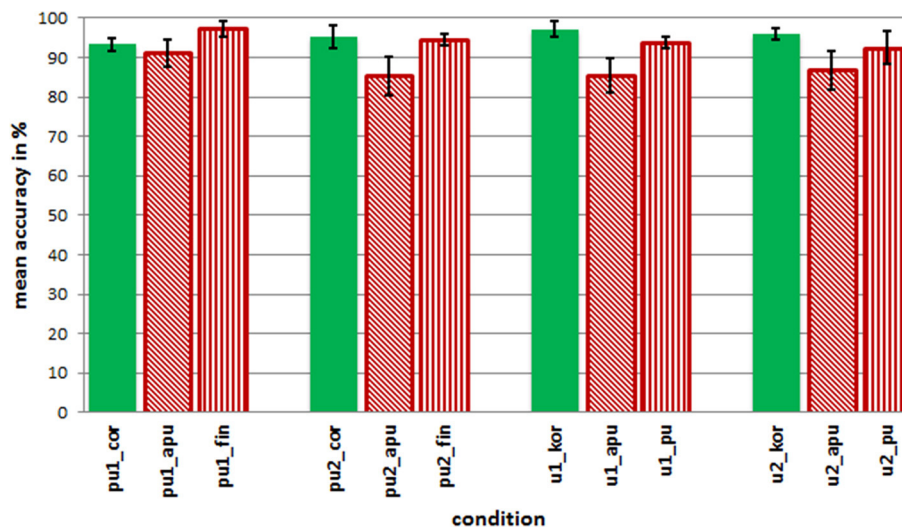


FIGURE 3 | Mean accuracy in percent for each word type and condition.

Table 4 | Statistical analyses of behavioral data.

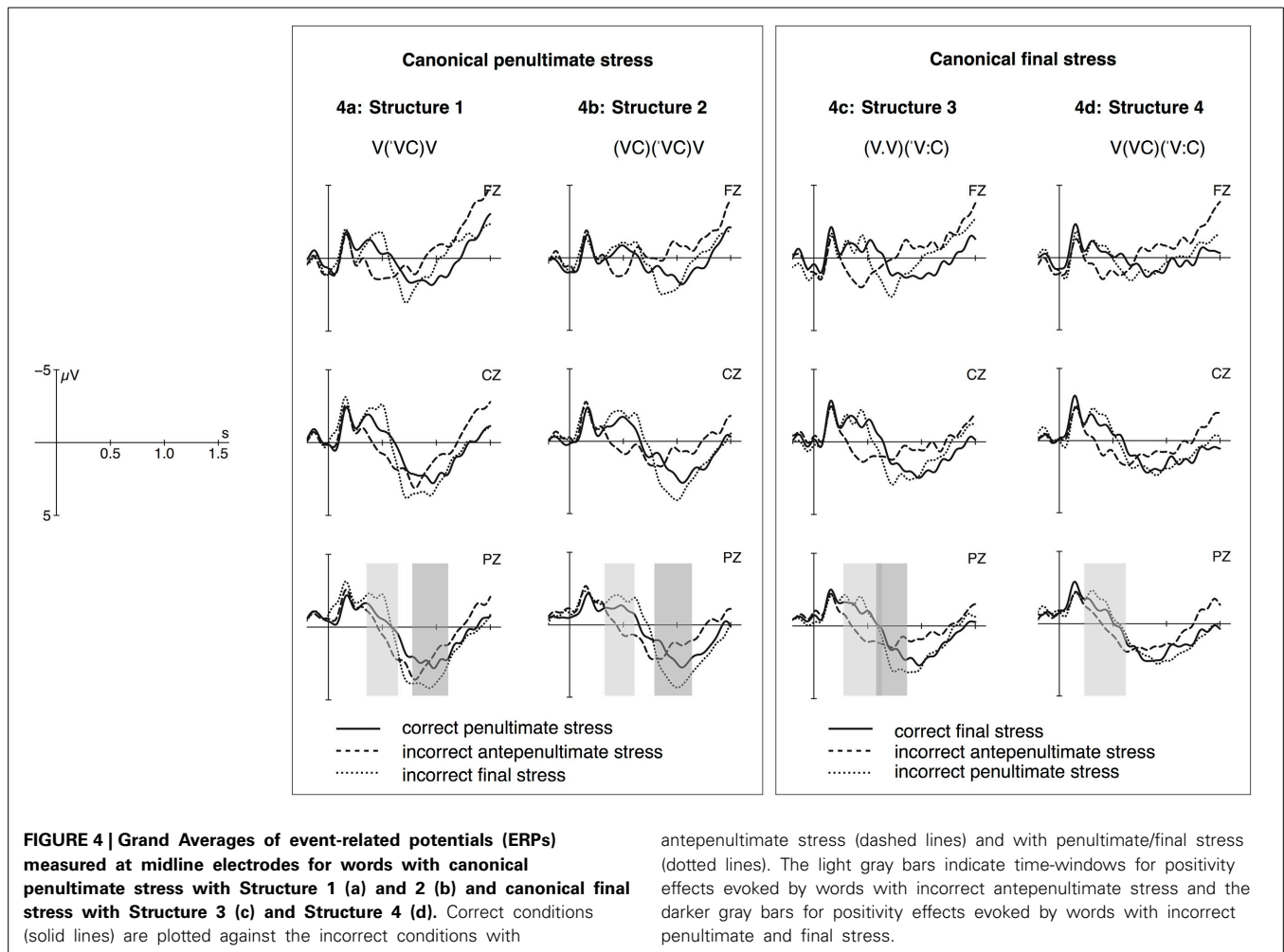
Analysis		Results
<b>ANOVA FOR WORDS WITH CANONICAL PENULTIMATE STRESS</b>		
FOOT STRUCTURE		$F_{(1, 19)} = 5.45; p < 0.04^*; \text{pes.223}$
STRESS POSITION		$F_{(2, 38)} = 7.59; p < 0.005^{**}; \text{pes.285}$
Interaction of FOOT STRUCTURE $\times$ stress position		$F_{(2, 38)} = 8.06; p < 0.004^{**}; \text{pes.298}$
<b>PAIRWISE <i>t</i>-TESTS</b>		
Word type 1 V(VC)V	Correct vs. antepenultimate stress	$t_{(19)} = -0.25; p > 0.80$
	Correct vs. final stress	$t_{(19)} = -2.26; p < 0.04$
	Antepenultimate vs. final stress	$t_{(19)} = -3.56; p < 0.003^{**}$
Word type 2 (VC)(V)V	Correct vs. antepenultimate stress	$t_{(19)} = -1.79; p > 0.08$
	Correct vs. final stress	$t_{(19)} = -2.38; p < 0.03$
	Antepenultimate vs. final stress	$t_{(19)} = -7.78; p < 0.001^{***}$
<b>ANOVA FOR WORDS WITH CANONICAL FINAL STRESS</b>		
FOOT STRUCTURE		$F_{(1, 19)} < 1$
STRESS POSITION		$F_{(2, 38)} = 5.36; p < 0.02^*; \text{pes} = 0.22$
Interaction of FOOT STRUCTURE $\times$ STRESS POSITION		$F_{(2, 38)} = 6.72; p < 0.004^{**}; \text{pes} = 0.261$
<b>PAIRWISE <i>t</i>-TESTS</b>		
Word type 3 V(V)(V:C)	Correct vs. antepenultimate stress	$t_{(19)} = -1.96; p > 0.06$
	Correct vs. penultimate stress	$t_{(19)} = 1.02; p > 0.32$
	Antepenultimate vs. penultimate stress	$t_{(19)} = -5.14; p < 0.001^{***}$
Word type 4 V(VC)(V:C)	Correct vs. antepenultimate stress	$t_{(19)} = -2.53; p < 0.02$
	Correct vs. penultimate stress	$t_{(19)} = -0.95; p > 0.35$
	Antepenultimate vs. penultimate stress	$t_{(19)} = -2.89; p < 0.01$

Repeated measures ANOVA for the two sets of words with canonical penultimate and final stress separately over the factors foot structure and stress position as well as pairwise *t*-tests for comparisons of correct with each of the two incorrect stress conditions and of both incorrect conditions. According to Bonferroni correction, the level of significance for paired *t*-tests is below 0.008. Significant results are indicated by  $< * >$ . Effect sizes are given by partial Eta-squared values (pes).

**Words with canonical final stress.** For violations with antepenultimate stress (dashed lines in **Figures 4C,D**), positivity effects occurred between 300 and 650 ms in both word types with canonical final stress. Repeated measures ANOVAs over the factors FOOT STRUCTURE, CORRECTNESS and REGION revealed a main effect for the factor CORRECTNESS and an interaction for CORRECTNESS  $\times$  REGION and CORRECTNESS  $\times$  FOOT STRUCTURE. *Post-hoc* analyses showed a difference between

correct final stress and incorrect antepenultimate stress in each REGION and each FOOT STRUCTURE (see **Table A2D**).

Violations with penultimate stress (dotted lines in **Figures 4C,D**) led to a negativity effect between 400 and 480 ms and to a positivity effect between 550 and 850 ms only in the context of word type 3 with the structure (V.V)(V:C), but not for word type 4 with a strong penultimate syllable V(VC)(V:C). For the negativity effect, a main effect for all



three factors but no interaction was found (see Table A2E), and for the positivity a main effect for the factors CORRECTNESS and REGION and an interaction between CORRECTNESS  $\times$  REGION as well as CORRECTNESS  $\times$  FOOT STRUCTURE. *Post-hoc* analyses suggest that an overall effect of CORRECTNESS is restricted to frontal regions only and that a difference between correct and incorrect penultimate stress occurs only for words of the structure (V.V) (V:C) (see Table A2F).

Figure 5 depicts mean amplitudes of respective peaks of positivity effects for correct and incorrect conditions measured at parietal electrodes (P3, Pz, P4). Except for incorrect penultimate stress in words with the structure 4 (V.(VC)(V:C); circled in Figure 5), the amplitude of positivity effects is significantly more pronounced in incorrect compared to correct conditions.

## DISCUSSION

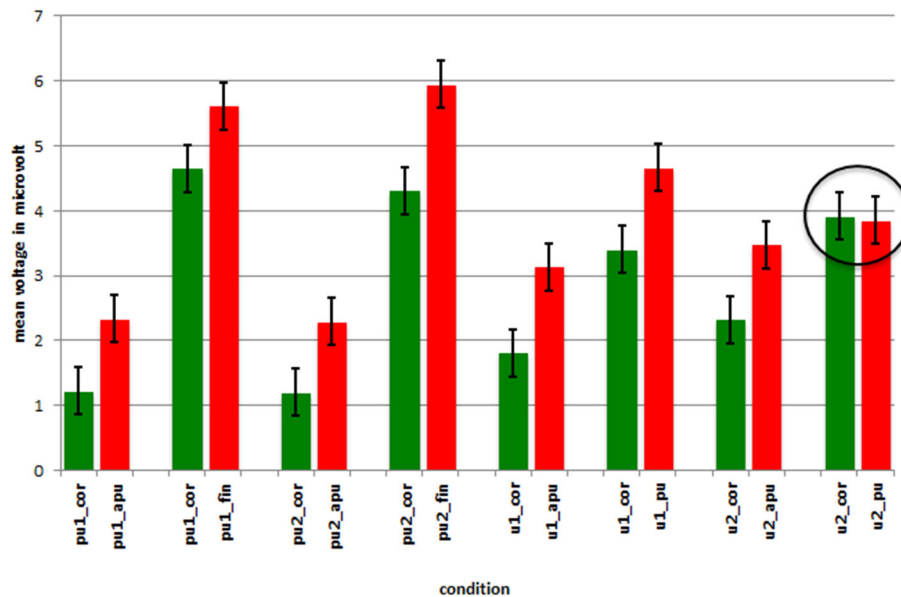
The current study aims at investigating whether speakers of Cairene Arabic are (like speakers of Turkish) partly insensitive to stress manipulations because stress in Cairene Arabic is predictable (as hypothesized in the Stress-“Deafness” account, i.e., Dupoux et al., 1997, 2001, 2008; Peperkamp and Dupoux,

2002), or whether the evaluation of stress differs between violations involving foot restructuring and those in which the prosodic structure is maintained.

In our ERP study utilizing a stress violation paradigm, violations of words with correct penultimate stress produced a positivity effect or a biphasic effect irrespective of prosodic structure: violations with antepenultimate stress evoked a positivity between 350 and 600 ms and violations with final stress a negativity between 400 and 550 ms and a positivity between 800 and 1150 ms. In contrast, for words with correct final stress asymmetrical results for different word structures are found: violations with antepenultimate stress evoked a positivity effect between 300 and 650 ms in both word types 3 and 4 and violations with penultimate stress a negativity between 400 and 480 ms, but a positivity only in word type 3 with the structure (V.V)(V:C) (between 550 and 850 ms).

We interpret the occurrence of positivity effects in different time-windows to reflect a task-related process that has been shown to reflect how easy it is for participants to decide how to classify a stress violation. We interpret these positivity effects as instances of the P3b family (Picton, 1992; Coulson et al., 1998; Nieuwenhuis et al., 2005) as found in previous similar experiments using the stress deviation paradigm (Domahs et al., 2008,





**FIGURE 5 | Mean amplitudes and standard errors in microvolt for positivity effects of correct and incorrect conditions measured at parietal electrodes.** The label pu1 denotes conditions of word type 1, pu2

conditions of word type 2, u1 conditions of word type 3 and u2 conditions of word type 4. The circle indicates the conditions for which the comparison was not significant.

2013). The P3b effect is known to reflect stimulus probability, saliency, and task relevance in diverse cognitive domains. According to Coulson et al. (1998), the P300 is an appropriate dependent variable to test the saliency of a given manipulation because the amplitude and the latency of the effect increase with the degree of anomaly. Thus, in the present study, violations evoking enhanced positivity effects can be regarded as less probable than violations with reduced effects. Overall, the positivity effects observed vary in latency, most likely due to the fact that the evaluation and decision-making process is dependent on the perception of a stressed syllable. Since strong syllables play a crucial role in the perception of stress patterns, the latency differences can be explained by varying positions of stressed syllables.

Generally, the findings of the experiment reported in Section ERP Experiment on Cairene Arabic show that stress deviations in Cairene Arabic words produce brain responses reflecting the participants' sensitivity to most violations. Their brain responses are similar to those obtained in previous experiments on German and Turkish. In the following, the results for specific word structures will be discussed in comparison to previous results.

#### ARE SPEAKERS OF CAIRENE ARABIC INSENSITIVE TO STRESS MANIPULATIONS?

In Section *Previous ERP Studies on Word Stress Processing*, results reported for speakers of Turkish showed that Turkish participants had difficulties judging incorrect stress patterns if the default stress pattern was applied to words with lexical stress, while violations of words with canonical default stress produced enhanced positivity effects (Domahs et al., 2013). This finding was interpreted as evidence for the insensitivity to the default stress pattern, and for the view that the processing of stress information

in Turkish mainly depends on the lexical status of stress (default vs. non-default stress). In Cairene Arabic, the position of word stress is also predictable though variable. In contrast to the Turkish default stress, stress in Cairene Arabic is not predictable by position but by structure. The behavioral data as well as the ERP data reported in Section *ERP Experiment on Cairene Arabic* suggest that speakers of Cairene Arabic are clearly sensitive to stress violations. In the behavioral data, correctly and incorrectly stressed words are accepted or rejected with an accuracy of more than 80%. Only violations involving incorrect antepenultimate stress are judged less accurately compared with other violations. However, this moderate difficulty is not reflected in ERPs in which violations with antepenultimate stress produced a positivity effect in each word type. In the study on Turkish, the condition with least accuracy in behavioral data did not produce a P300 effect.

In words with the structure 4 [V(VC)(V:C); e.g., ki.(ris)('ta:l)] with canonical final stress in Cairene Arabic, a lack of a positivity effect occurs for incorrect penultimate stress. We argue that the absence of a positivity cannot be explained by the factor predictability in the sense that penultimate stress is the default stress. In words like ki.(ris)('ta:l) final stress is the only predicted stress pattern. The most reasonable explanation is related to the metrical structure of phonological words in Cairene Arabic as discussed in the following section.

#### THE ROLE OF THE METRICAL STRUCTURE IN STRESS PROCESSING

Related to the findings on German word stress processing (as summarized in Section *Previous ERP Studies on Word Stress Processing*), the second question was to test whether word stress processing in Cairene Arabic is guided by the internal foot structure of phonological words. In Table 5, the structures of correct forms are compared with those of incorrect forms.

In **Table 5** it can be seen that in words exhibiting more than one foot (structures 2–4) violations occur that do not involve restructuring of feet, i.e., neither regrouping of syllables into feet nor creating feet from unparsed syllables. The results from the experiment on German (Domahs et al., 2008; see Section German) suggested a qualitative distinction between violations with stress realized on the head syllable of a weak foot and violations with stress on a weak or unparsed syllable. Thus, in German it was possible to identify indirectly which syllables are capable of bearing stress and which are not via the occurrence of P300 effects. With respect to the experiment on Cairene Arabic, it was expected that violations with stress on the head syllable of a weak foot are more difficult to classify as violation than violations involving changed structure, the latter ones leading to a P300 component. From the occurrences of P300 effects (**Table 5**) in the experiment on Cairene Arabic it seems that our hypothesis is not borne out in all cases: A lack of a P300 effect was obtained only for violations with penultimate stress when the structure was preserved (see final row in **Table 5**), but violations involving antepenultimate stress produce P300 effects in each word type, although in words with structure 2 and 3 such violations maintain the foot structure.

The question arises whether the effect patterns found in the study on Cairene Arabic can be interpreted along the same lines as the results found for German. We suggest that structure plays a role in Cairene Arabic stress processing when certain conditions are met: first the structure is maintained and second the incorrect stress pattern involved is a likely pattern in terms of frequency. Thus, we hypothesize that metrical structure is not the only factor influencing stress perception, but also the frequency asymmetries between different stress patterns. To strengthen this hypothesis we report the results of a frequency count on stress patterns in loan words.

An analysis of stress patterns in loan words in Cairene Arabic by El Shanawany (2013) showed that irrespective of the stress position in the source language, stress is assigned along the principles also suggested for native words of Cairene Arabic and is predictable by syllable quantity and position of (the head of) the final foot in phonological words. The corpus analyzed consisted

of loan words because the trisyllabic stimuli presented in the ERP study are predominantly loans. Out of 286 types of bi-, tri-, and quadrisyllabic words, 57% exhibit final stress, 39% penultimate stress, and only 4% antepenultimate stress. Since native words of Cairene Arabic consist of higher proportions of mono- and bisyllabic words than loan words, the proportion of words with antepenultimate stress among native words can be expected to be even lower than 4%. Antepenultimate stress occurs only in words with three light syllables, a rare configuration. This corpus analysis demonstrates that final feet are more likely to be aligned with the right than with the left edge of phonological words. In this respect, Cairene Arabic differs from German for which it is postulated that the final foot within words is strong but which exhibits many exceptions with stress on non-final feet [e.g., 69% of existing words of the structure (V.V)(VC); see Janssen (2003)]. The positivity effect in words with incorrectly stressed head syllables in antepenultimate position (structures 2 and 3) indicate that such violations are clearly identified as deviating patterns though the participants were less accurate in explicitly judging them as incorrect compared to other violations. This discrepancy between behavioral and electrophysiological data suggests that the P300 effect not simply reflects the explicit judgment but rather the implicit evaluation of the likeliness of an event. One potential explanation for the occurrence of the P300 effect in words that preserve the prosodic structure could be that antepenultimate stress involving left aligned strong feet occur only rarely in Cairene Arabic and could therefore be classified as exceptional. In principle, the sensitivity to exceptional, less frequent stress patterns was also demonstrated in the study on Turkish word stress, in which only exceptional incorrect stress patterns led to P3 effects. Antepenultimate stress in Cairene Arabic is not exceptional in the sense that it is not derived by foot structure, but rather in terms of stress pattern frequency: only a few words consist of a sequence of three light syllables.

Taken together, the occurrence or absence of P3 effects in Cairene Arabic seems to be guided by the metrical structure and by the frequency distribution of the different stress positions, i.e., whether a certain pattern is exceptional or not. Therefore, we suggest that the participants' performance and sensitivity to word stress violations lie in between those observed for Turkish and German participants. Comparable to Turkish, exceptional stress patterns evoke a P3 effect when used incorrectly, and comparable to German, metrical structure plays a role. In contrast to Turkish, Cairene Arabic exhibits no default pattern, and in contrast to German word stress shows a stronger orientation toward the right edge of words.

#### NEGATIVITY EFFECT: ERROR-DETECTION MECHANISM OR VIOLATION OF LEXICAL EXPECTANCY?

In Section Results, it was reported that violations involving penultimate and final stress evoked a biphasic ERP pattern. The discussion so far has mainly focused on the interpretation of the positivity effect. As regards the negativity effect in similar experiments, different interpretations have been proposed in the literature. In the study on German word stress processing (Domahs et al., 2008), an extended more fronto-centrally distributed negativity was found which was interpreted as an instance of a

**Table 5 | Overview of metrical structures in correctly and incorrectly stressed forms and the occurrence of P3 effects as reflections of task-specific evaluation-to-expectation processes.**

Canonical Structure	Structure violation	Restructuring of feet	Occurrence of P3
1 V('VC)V	*('V)(VC)V	Yes	Yes
	*V(VC)('V)	Yes	Yes
2 (VC)('VC)V	*('VC)(VC)V	No	Yes
	*('VC)(VC)('V)	Yes	Yes
3 (V.V)('VVC)	*('V.V)(VVC)	No	Yes
	*V('V)(VVC)	Yes	Yes
4 V(VC)('VVC)	*('V)(VC)(VVC)	Yes	Yes
	*V('VC)(VVC)	No	No

contingent negative variation (CNV; according to Rugg, 1984) to reflect the detection of a pitch-contour violation when a de-stressed initial syllable was encountered that did not provide sufficient information to judge such a form as incorrect. The judgment requires the detection of a stressed syllable (Domahs et al., 2008, 2013). In the present experiment on Cairene Arabic, however, the occurrences of negativity effects do not seem to mirror the perception of de-stressing and the prolonged activation of the phonological form in the working memory. The negativity effects occur for violations with penultimate and final stress, and in both cases the curve is not flat and extended over more than 400 ms (slow wave) but peaks at around 400–550 ms (see Figure 4).

In the study on Turkish word stress processing (Domahs et al., 2013), a centro-parietal negativity effect between 500 and 750 ms was obtained for violations with the default pattern (= final stress) replacing lexical penultimate stress. The effect was interpreted as belonging to the N400 family. For Turkish, it was assumed that exceptional stress on the penultimate or antepenultimate syllable has to be lexically specified in the phonological representations of words. If the lexical specification is not realized, the violation of the stress expectation leads to an N400 effect.

For Cairene Arabic, in contrast, it is not very likely that the negativity effects reflect deviations from lexical expectations. There are no indications that stress positions need to be lexically specified in Cairene Arabic. Furthermore, the components occur earlier than in the Turkish experiment (between 400 and 480 ms or 400–550 ms instead of 500–750 ms in Turkish). In previous studies on metrical processing (e.g., Koelsch et al., 2000; Rothermich et al., 2010), negativity effects were observed that have been proposed to indicate the general detection of deviations in metrical regularity or expectation. This component which has been described with different distributions (either lateralized or not, more frontally or broadly) and which has therefore been labeled differently, can be roughly summarized as an error detection component. It is suggested here that the present negativity effects represent an error detection mechanism, which is independent from lexical processing but related to metrical deviations. This component is independent from the occurrence of the later P3 effect as becomes evident for violations with penultimate stress in words with heavy penults [ $*V(VC)(V:C$  e.g.,  $*ki(ris)(ta:l)$ ]. Thus, participants detect the metrical error, but in the evaluation process such violations are difficult to categorize as an unlikely form.

## CONCLUSION

The present behavioral and electrophysiological results on stress perception in Cairene Arabic show that speakers of this language are sensitive to stress information because they perform accurately in a stress evaluation task and produce ERP components indicating their ability to evaluate and categorize the likeliness of a certain stress pattern. Thus, psycholinguistic accounts of stress perception like the *Stress “Deafness”* account (i.e., Dupoux et al., 1997, 2001, 2008; Peperkamp and Dupoux, 2002), which assume that speakers of a language with predictable stress have difficulties identifying stress information, cannot explain the effect patterns we found.

Rather, our data support linguistic theories proposed for the Cairene Arabic word stress system as outlined in Section Metrical Properties of Cairene Arabic. In particular it was shown that prosodic structure, and metrical feet in particular, determines stress perception. This was evident for the processing of incorrect penultimate stress evoking a late positivity effect only if a light penult was stressed, but not when it was heavy. However, this structure effects cannot be generalized to incorrect antepenultimate stress which was easily categorized as unlikely irrespective of weight and its position within feet. To account for this result, it has been suggested that the frequency of stress patterns influences the processing of word stress in Cairene Arabic as a second factor. This hypothesis is supported by a corpus analysis of loan words. Effects of stress perception in Cairene Arabic lie therefore in between those obtained for German and Turkish.

Together with previous findings on stress perception in German and Turkish the present data complement the results by Dupoux, Peperkamp and colleagues that stress sensitivity is a function of predictability of stress. Our results suggest that the metrical structure in foot-based systems (i.e., German, Cairene Arabic), the lexical status of stress patterns in languages with default and lexical (exceptional) stress (i.e., Turkish), and the frequency of certain patterns also influences stress perception.

## ACKNOWLEDGMENTS

The research presented here was funded by the German Science Foundation grant WI853/7-2 to Richard Wiese and Ulrike Domahs. We are grateful to the participants from Cairo, Egypt, and to Karen Bohn and Janine Kleinhans for their help in collecting the data. Last but not least we would like to thank the reviewers for their valuable comments.

## REFERENCES

- Coulson, S., King, J. W., and Kutas, M. (1998). Expect the unexpected: event-related brain responses to morphosyntactic violations. *Lang. Cogn. Process.* 13, 21–58. doi: 10.1080/016909698386582
- Domahs, U., Genc, S., Knaus, J., Wiese, R., and Kabak, B. (2013). Processing (un-)predictable word stress: ERP evidence from Turkish. *Lang. Cogn. Process.* 28, 335–354. doi: 10.1080/01690965.2011.634590
- Domahs, U., Plag, I., and Carroll, R. (2014). Word stress assignment in German, English and Dutch: quantity-sensitivity and extrametricality revisited. *J. Comp. Ger. Linguist.* 17, 59–96. doi: 10.1007/s10828-014-9063-9
- Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I. D., and Schlewsky, M. (2008). The processing of German word stress: evidence for the prosodic hierarchy. *Phonology* 25, 1–36. doi: 10.1017/S0952675708001383
- Dupoux, E., Pallier, C., Sebastián, N., and Mehler, J. (1997). A destressing “deafness” in French? *J. Mem. Lang.* 36, 406–421. doi: 10.1006/jmla.1996.2500
- Dupoux, E., Peperkamp, S., and Sebastián-Gallés, N. (2001). A robust method to study stress “deafness.” *J. Acoust. Soc. Am.* 110, 1606–1618. doi: 10.1121/1.1380437
- Dupoux, E., Peperkamp, S., and Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress ‘deafness’ simultaneous French-Spanish bilinguals. *Cognition* 114, 266–275. doi: 10.1016/j.cognition.2009.10.001
- Dupoux, E., Sebastián-Gallés, N., Navarrete, E., and Peperkamp, S. (2008). Persistent stress ‘deafness’: the case of French learners of Spanish. *Cognition* 106, 682–706. doi: 10.1016/j.cognition.2007.04.001
- Eisenberg, P. (1991). Syllabische Struktur und Wortakzent: Prinzipien der Prosodik deutscher Wörter. *Z. Sprachwissenschaft* 10, 37–64. doi: 10.1515/zfsw.1991.10.1.37
- El Shanawany, H. (2013). *Der Kairo-Arabisches Wortakzent im Vergleich Zum Deutschen: Eine EEG-Untersuchung [A Comparison Between Cairene-Arabic and German Word Stress: an ERP Study]*. Ph.D. dissertation, Philipps-Universität

- Marburg. Available online at: <http://archiv.ub.uni-marburg.de/diss/z2014/0068/>
- Féry, C. (1998). German word stress in optimality theory. *J. Comp. Ger. Linguist.* 2, 101–142. doi: 10.1023/A:1009883701003
- Göksel, A., and Kerslake, C. (2005). *Turkish: A Comprehensive Grammar*. London/New York: Routledge.
- Harrell, R. (1960). “A linguistic analysis of egyptian radio arabic,” in *Contributions to Arabic Linguistics*, ed C. A. Ferguson (Cambridge, MA: Harvard University Press), 3–77.
- Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.
- Hulst van der, H., and Hellmuth, S. (2010). “Word accent systems in the middle east,” in *A Survey of Word Accentual Patterns in the Languages of the World, chap. 11*, eds H. Hulst van der, R. Goedemans, and E. van Zanten (Berlin; New York: Mouton de Gruyter), 615–646.
- Hyman, L. (1985). *A Theory of Phonological Weight*. Dordrecht: Foris.
- Inkelas, S. (1999). “Exceptional stress-attracting suffixes in Turkish: representations versus the grammar,” in *The Prosody-Morphology Interface*, eds R. Kager, H. Hulst van der, and W. Zonneveld (Cambridge: Cambridge University Press), 134–187.
- Inkelas, S., and Orgun, C. O. (2003). Turkish stress: a review. *Phonology* 20, 139–161. doi: 10.1017/S0952675703004482
- Janssen, U. (2003). *Untersuchungen Zum Wortakzent im Deutschen und Niederländischen*. Düsseldorf: Heinrich-Heine-Universität Düsseldorf.
- Kabak, B., and Vogel, I. (2001). The phonological word and stress assignment in Turkish. *Phonology* 18, 315–360. doi: 10.1017/S0952675701004201
- Kabak, B., and Vogel, I. (2011). “Exceptions to stress and harmony: cophonologies or prespecification?” in *Expecting the Unexpected: Exceptions in Grammar*, eds H. J. Simon and H. Wiese (Berlin/New York: Mouton de Gruyter), 59–94.
- Kaisse, E. (1985). “Some theoretical consequences of stress rules in Turkish,” in *Papers from the General Session of the 21st Regional Meeting*, eds W. Eilfort, P. Kroeber, and K. Peterson (Chicago, IL: Chicago Linguistic Society), 199–209.
- Knaus, J., and Domahs, U. (2009). Experimental evidence for optimal and minimal metrical structure of German word prosody. *Lingua* 119, 1396–1413. doi: 10.1016/j.lingua.2008.04.002
- Knaus, J., Wiese, R., and Janssen, U. (2007). “The processing of word stress: EEG studies on task-related components,” in *Proceedings of the 16th International Conference of Phonetic Sciences* (Saarbrücken), 709–712.
- Koelsch, S., Gunter, T., Friederici, A. D., and Schröger, E. (2000). Brain indices of music processing: “Nonmusicians” are musical. *J. Cogn. Neurosci.* 12, 520–541. doi: 10.1162/089892900562183
- Kornfilt, J. (1997). *Turkish*. London; New York: Routledge.
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning of the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123
- Lewis, G. (1967/2000). *Turkish Grammar*. Oxford: Oxford University Press.
- Marie, C., Magne, C., and Besson, M. (2011). Musicians and the metric structure of words. *J. Cogn. Neurosci.* 23, 294–305. doi: 10.1162/jocn.2010.21413
- McCarthy, J. (1979). *Formal Problems in Semitic Phonology and Morphology*. Ph.D. dissertation, MIT.
- Mitchell, T. F. (1960). Prominence and syllabification in Arabic. *Bull. Sch. Orient. Afr. Stud.* 2, 369–389. doi: 10.1017/S0041977X00149997
- Nieuwenhuis, S., Aston-Jones, G., and Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus–norepinephrine system. *Psychol. Bull.* 131, 510–532. doi: 10.1037/0033-2909.131.4.510
- Peperkamp, S., and Dupoux, E. (2002). “A typological study of stress ‘deafness,’” in *Laboratory Phonology 7*, eds C. Gussenhoven and N. Warner (Berlin: Mouton de Gruyter), 203–240.
- Peperkamp, S., Vendelin, I., and Dupoux, E. (2010). Perception of predictable stress: a cross-linguistic investigation. *J. Phon.* 38, 422–430. doi: 10.1016/j.wocn.2010.04.001
- Picton, T. W. (1992). The P300 wave of the human event-related brain potential. *J. Clin. Neurophysiol.* 9, 456–479. doi: 10.1097/00004691-199210000-00002
- Rothermich, K., Schmidt-Kassow, M., Schwartz, M., and Kotz, S. A. (2010). Event-related potential responses to metric violations: rules versus meaning. *Neuroreport* 21, 580–584. doi: 10.1097/WNR.0b013e32833a7da7
- Röttger, T. B., Domahs, U., Grande, M., and Domahs, F. (2012). Structural factors affecting the assignment of word stress in German. *J. Ger. Linguist.* 24, 53–94. doi: 10.1017/S1470542711000262
- Rugg, M. D. (1984). Event-related potentials in phonological matching tasks. *Brain Lang.* 23, 225–240. doi: 10.1016/0093-934X(84)90065-8
- Rugg, M. D., and Nagy, M. E. (1989). Event-related potentials and recognition memory for words. *Electroencephalogr. Clin. Neurophysiol.* 72, 395–406.
- Schmidt-Kassow, M., Roncaglia-Denissen, M. P., and Kotz, S. (2011a). Why pitch sensitivity matters: event-related potential evidence of metrical and syntactic violation detection among Spanish late learners of German. *Front. Psychol.* 2:131. doi: 10.3389/fpsyg.2011.00131
- Schmidt-Kassow, M., Rothermich, K., Schwartz, M., and Kotz, S. (2011b). Did you get the beat? Late proficient French-German learners extract strong-weak patterns in tonal but not in linguistic sequences. *Neuroimage* 54, 568–576. doi: 10.1016/j.neuroimage.2010.07.062
- Sezer, E. (1981). On non-final stress in Turkish. *J. Turkish Stud.* 5, 61–69.
- Watson, J. C. E. (2002). *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.
- Wiese, R. (1996). *The Phonology of German*. Oxford: Oxford University Press.
- Woidich, M. (2006). *Das Kairenisch-Arabishe. Eine Grammatik*. Wiesbaden: Harrassowitz Verlag.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 February 2014; accepted: 23 September 2014; published online: 21 October 2014.

Citation: Domahs U, Knaus JA, El Shanawany H and Wiese R (2014) The role of predictability and structure in word stress processing: an ERP study on Cairene Arabic and a cross-linguistic comparison. *Front. Psychol.* 5:1151. doi: 10.3389/fpsyg.2014.01151  
This article was submitted to Language Sciences, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Domahs, Knaus, El Shanawany and Wiese. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDIX

Table A1 | List of critical items.

Structure	Stimuli	Transcription	English translation
1. V(VX)V	فانيلا	va.nil.ʃa	vanilla
	أجندة	ʔa.ʒin.da	agenda
	كاريزما	ka.riz.ma	charisma
	شهادة	ʃi.ha:.da	certification
	حضنة	ha.ɗa:.na	kindergarten
	سفارة	si.ʃa:.ra	embassy
	جزيرة	gi.zi:.ra	island
	عروسة	ʕa.ro:.sa	bride
	دقيقة	di.ʔi:.ʔa	minute
	بيجامة	bi.ʒa:.ma	pajamas
	دراسة	di.ra:.sa	study
	طريقة	Tʔa..ri.ʔa	method
	تفاهم	ta.ʃa:hom	agreement
	نموذج	na.mo:.zag	pattern
	كابينة	ka.bi:.na	cabin
2. (VC)(VC)V	مستشفى	mus.taʃ.ʃa	hospital
	شخصية	ʃax.ʃij.ʃa	personality
	هستيريا	his.tir.ʃa	hysteria
	إستيريو	ʔis.tir.ju	stereo
	بكتيريا	bak.tir.ʃa	bacillus
	بلياردو	bil.jar.du	billard
	شمبانيا	ʃam.ban.ʃa	champagne
	شمبانزي	ʃam.ban.zi	chimp
	هامبورجر	ham.bur.gar	hamburger
	تأشيرة	Taʔ.ʃi:.ra	visa
	بسكوطة	bas.ko:.ta	cookie
	فترينا	vat.ri:.na	cabinet
	أنتيكة	ʔan.ti:.ka	antiquity
	برنامج	bir.na:.mig	programm
	دفاية	daf.ʃa:.ja	heating
3. (V.V)(V:C)	مصاريف	ma.sʔa.ri:f	costs
	بلاتين	bi.la.ti:n	platinum
	فيتامين	vi.ta.mi:n	vitamine
	كتالوج	ka.ta.lo:g	catalog
	كوافير	ko.wa.fe:r	hairdresser
	ليموزين	li.mo.zi:n	limousine
	مايونيز	ma.jo.ne:z	mayonnaise
	مانيكان	ma.ni.ka:n	model
	مانيكير	ma.ni.ke:r	nail polish
	مونولوج	mo.no.lo:g	monologue
	سمينار	si.mi.na:r	seminar
	تليفون	ti.li.fo:n	telephone
	جيلاتين	ʒi.la.ti:n	gelatin
	أناناس	ʔa.na.na:s	pineapple
	إتيكيت	ʔe.ti.ke:t	lable
4. V(VC)(V:C)	كريستال	ki.ris.ta:l	crystal
	برستيج	bi.ris.ti:ʒ	prestige
	ديكولتيه	di.kol.te:h	décolleté
	جليسرين	gi.lis.ri:n	glycerin
	كوميديان	ko.mid.ʃa:n	comedian
	كومسيون	ko.mis.ʃo:n	commettee
	ماجستير	ma.ʒis.te:r	master
	سكرتير	se.kir.te:r	secretary
	تليسكوب	te.lis.ko:b	telescope
	أوكلزيون	ʔo.kaz.ʃo:n	sale
	بدنجان	bi.din.ga:n	aubergine
	موريستان	mo.ris.ta:n	madhouse
	حيكمدار	hi.kim.da:r	police grade (dated)
	ترايزين	ta.rab.zi:n	banister
	أسانسير	ʔa.san.se:r	lift

## APPENDIX 2: OVERVIEW OVER STATISTICAL RESULTS

Generalized repeated measures ANOVAs of mean voltage changes over the factors Foot Structure (the two structures per canonical stress pattern), Correctness (correct vs. incorrect stress condition) and Region (frontal, central, parietal electrodes) and *post-hoc* analyses of interactions with Bonferroni correction. Effect sizes are given in generalized eta-squared values ( $\eta^2$ ). Tables A2A–A2F provide summaries for separate comparisons of correct and incorrect conditions. U is the abbreviation for final stress, PU for penultimate stress and APU for antepenultimate stress. 1 and 2 refers to different foot structures.



**Table A2 | Conditions and material.**

<b>(A) PU1/PU2 correct vs. incorrect APU stress/positivity in time-window 350 to 600 ms</b>					
<b>PU1: CV(CVC)CV</b>					
<b>PU2: (CVC)(CVC)CV</b>					
	<b>DFn</b>	<b>DFd</b>	<b>F</b>	<b>p</b>	<b>ges</b>
<b>Main effects</b>					
Region	2	38	6.273	<b>0.004</b>	0.033
Correctness	1	19	23.785	<b>0.000</b>	0.195
Foot structure	1	19	0.113	0.740	0.001
<b>Interactions</b>					
Region:Correctness	2	38	5.596	<b>0.007</b>	0.005
Region:Foot structure	2	38	0.332	0.720	0.000
Correctness:Foot structure	1	19	0.090	0.768	0.000
Region:Correctness:Foot structure	2	38	5.737	<b>0.007</b>	0.005
<b>Post-Hoc Analyses (p-Values Bonferroni-Corrected)</b>					
Region:Correctness					
Frontal	1	19	28.27	<b>0.000</b>	0.286
Central	1	19	26.08	<b>0.000</b>	0.288
Parietal	1	19	13.02	<b>0.006</b>	0.288
Region:Correctness:Foot structure					
Foot Structure 1 frontal	1	19	39.04	<b>0.000</b>	0.327
Foot Structure 1 central	1	19	18.26	<b>0.001</b>	0.18
Foot Structure 1 parietal	1	19	8.003	<b>0.032</b>	0.074
Foot Structure 2 frontal	1	19	10.58	<b>0.013</b>	0.158
Foot Structure 2 central	1	19	23.01	<b>0.000</b>	0.33
Foot Structure 2 parietal	1	19	12.2	<b>0.007</b>	0.197
<b>(B) PU1/PU2 correct vs. incorrect U stress/negativity in time-window 400 to 550 ms</b>					
<b>PU1: CV(CVC)CV</b>					
<b>PU2: (CVC)(CVC)CV</b>					
	<b>DFn</b>	<b>DFd</b>	<b>F</b>	<b>p</b>	<b>ges</b>
<b>Main effect</b>					
Region	2	38	6.540	<b>0.004</b>	0.040
Correctness	1	19	8.998	<b>0.007</b>	0.083
Foot structure	1	19	0.053	0.821	0.000
<b>Interactions</b>					
Region:Correctness	2	38	2.947	0.065	0.003
Region:Foot structure	2	38	5.072	<b>0.011</b>	0.007
Correctness:Foot structure	1	19	3.895	0.063	0.018
Region:Correctness:Foot structure	2	38	0.482	0.621	0.000
<b>Post-Hoc Analyses (p-Values Bonferroni-Corrected)</b>					
Region:Foot structure					
Frontal	1	19	2.066	0.501	0.023
Central	1	19	<1		0
Partietal	1	19	<1		0.006

(Continued)

Table A2 | Continued

<b>(C) PU1/PU2 correct vs. incorrect U stress/positivity in time-window 800 to 1150 ms</b>					
<b>PU1: CV(CVC)CV</b>					
<b>PU2: (CVC)(CVC)CV</b>					
	<b>DFn</b>	<b>DFd</b>	<b>F</b>	<b>p</b>	<b>ges</b>
<b>Main effects</b>					
Region	2	38	27.595	<b>0.000</b>	0.167
Correctness	1	19	9.265	<b>0.007</b>	0.043
Foot structure	1	19	0.169	0.686	0.001
<b>Interactions</b>					
Region:Correctness	2	38	35.409	<b>0.000</b>	0.031
Region:Foot structure	2	38	4.527	<b>0.017</b>	0.003
Correctness:Foot structure	1	19	3.057	0.097	0.012
Region:Correctness:Foot structure	2	38	7.326	<b>0.002</b>	0.004
<b>Post-Hoc Analyses (p-Values Bonferroni-Corrected)</b>					
Region:Correctness:Word structure					
Foot Structure 1 frontal	1	19	3.47	0.234	0.05
Foot Structure 1 central	1	19	3.724	0.206	0.037
Foot Structure 1 parietal	1	19	15.4	<b>0.003</b>	0.136
Foot Structure 2 frontal	1	19	3.063	0.289	0.029
Foot Structure 2 central	1	19	12.3	<b>0.007</b>	0.092
Foot Structure 2 parietal	1	19	18.1	<b>0.001</b>	0.135
<b>(D) U1/U2 correct vs. incorrect APU stress/positivity in time-window 300 to 650 ms</b>					
<b>U1: (CV.CV)(CVVC)</b>					
<b>U2: CV(CVC)(CVVC)</b>					
	<b>DFn</b>	<b>DFd</b>	<b>F</b>	<b>p</b>	<b>ges</b>
<b>Main effects</b>					
Region	2	38	2.618	0.086	0.013
Correctness	1	19	70.228	<b>0.000</b>	0.270
Foot structure	1	19	0.955	0.341	0.004
<b>Interactions</b>					
Region:Correctness	2	38	7.630	<b>0.002</b>	0.007
Region:Foot structure	2	38	2.048	0.143	0.002
Correctness:Structure	1	19	4.480	<b>0.048</b>	0.012
Region:Correctnes:Foot structure	2	38	0.140	0.870	0.000
<b>Post-Hoc Analyses (p-Values Bonferroni-Corrected)</b>					
Region:Correctness					
Frontal	1	19	59.07	<b>0.000</b>	0.359
Central	1	19	79	<b>0.000</b>	0.368
Parietal	1	19	45.26	<b>0.000</b>	0.205
Correctness:Foot Structure					
Foot Structure 1	1	19	74.58	<b>0.000</b>	0.465
Foot Structure 2	1	19	27.41	<b>0.000</b>	0.182
<b>(E) U1/U2 correct vs. incorrect APU stress/negativity in time-window 400 to 480 ms</b>					
<b>U1: (CV.CV)(CVVC)</b>					
<b>U2: CV(CVC)(CVVC)</b>					
	<b>DFn</b>	<b>DFd</b>	<b>F</b>	<b>p</b>	<b>ges</b>
<b>Main effects</b>					
Region	2	38	8.026	<b>0.001</b>	0.034
Correctness	1	19	6.957	<b>0.016</b>	0.060
Foot structure	1	19	5.002	<b>0.038</b>	0.013

(Continued)

**Table A2 | Continued**

<b>(E) U1/U2 correct vs. incorrect APU stress/negativity in time-window 400 to 480 ms</b>					
<b>U1: (CV.CV)(CVVC)</b>					
<b>U2: CV(CVC)(CVVC)</b>					
<b>Interactions</b>					
Region:Correctness	2	38	0.479	0.623	0.001
Region:Foot structure	2	38	0.048	0.953	0.000
Correctness:Foot structure	1	19	2.082	0.165	0.006
Region:Correctness:Foot structure	2	38	1.192	0.315	0.000
<b>(F) U1/U2 correct vs. incorrect PU stress/positivity in time-window 550 to 850 ms</b>					
<b>U1: (CV.CV)(CVVC)</b>					
<b>U2: CV(CVC)(CVVC)</b>					
	<b>DFn</b>	<b>DFd</b>	<b>F</b>	<b>p</b>	<b>ges</b>
<b>Main effects</b>					
Region	2	38	4.073	<b>0.025</b>	0.027
Correctness	1	19	4.552	<b>0.046</b>	0.041
Foot structure	1	19	0.094	0.763	0.000
<b>Interactions</b>					
Region:Correctness	2	38	4.250	<b>0.022</b>	0.005
Region:Foot structure	2	38	2.567	0.090	0.003
Correctness:Foot structure	1	19	7.909	<b>0.011</b>	0.011
Region:Correctness:Foot structure	2	38	2.749	0.077	0.001
<b>Post-Hoc Analyses (p-Values Bonferroni-Corrected)</b>					
Region:Correctness					
Frontal	1	19	6.423	0.061	0.107
Central	1	19	4.325	0.154	0.05
Parietal	1	19	1.673	0.634	0.012
Correctness:Word structure					
Foot Structure 1	1	19	8.73	<b>0.016</b>	0.112
Foot Structure 2	1	19	1.031	0.645	0.012



# Evidence for the role of German final devoicing in pre-attentive speech processing: a mismatch negativity study

Hubert Truckenbrodt<sup>1,2\*</sup>, Johanna Steinberg<sup>3</sup>, Thomas K. Jacobsen<sup>4</sup> and Thomas Jacobsen<sup>4</sup>

<sup>1</sup> Centre for General Linguistics, Berlin, Germany

<sup>2</sup> Institut für deutsche Sprache und Linguistik, Humboldt University, Berlin, Germany

<sup>3</sup> Department of Psychology, University of Leipzig, Leipzig, Germany

<sup>4</sup> Experimental Psychology Unit, Helmut Schmidt University/University of the Federal Armed Forces Hamburg, Hamburg, Germany

## Edited by:

Ulrike Domahs, University of Marburg, Germany

## Reviewed by:

Carsten Eulitz, University of Konstanz, Germany

Matthew Winn, University of Wisconsin–Madison, USA

Burkhard Maess, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

## \*Correspondence:

Hubert Truckenbrodt, ZAS, Schützenstraße 18, 10117 Berlin, Germany  
e-mail: truckenbrodt@zas.gwz-berlin.de

Results of a mismatch negativity experiment are reported in which the pre-attentive relevance of the German phonological alternation of final devoicing (FD) is shown in two ways. The experiment employs pseudowords. (1) A deviant [vus] paired with standard /vuzə/ did not show a mismatch effect for the voicing change in /z/ versus [s] because the two can be related by FD. When standard and deviant were reversed, the two could not be related by FD and a mismatch effect for the voicing difference occurred. (2) An ill-formed deviant that violates FD, \*[vuz], triggered mismatch effects that were plausibly attributed to its ill-formedness. The results show that a syllable-related process like FD is already taken into account by the processing system in early pre-attentive processing.

**Keywords:** mismatch negativity (MMN), event-related potentials (ERP), phonological rules, final devoicing, phonotactics, German, pre-attentive processing

## INTRODUCTION

### NEURAL PROCESSING AND MISMATCH NEGATIVITY

Electrophysiological methods like the electroencephalogram (EEG) and the magnetoencephalogram (MEG) provide the possibility to obtain online insight into the perceptual process. This includes the pre-conscious and pre-attentive or automatic stages of processing. There is a sequence of positive–negative–positive deflections in the event-related potential (ERP), of which the first negative deflection (N100) typically peaks around 100 ms after the occurrence of a transient sound like an isolated vowel. The N100 has been found for speech sounds and for non-speech sounds. For speech sounds, the processing in this early stage shows, for one thing, characteristics of acoustic processing that are independent of phonological categories (e.g., Sharma and Dorman, 2000). At the same time, a number of studies have demonstrated the effect of phonological categories in this early stage: acoustically equidistant stimuli cluster along phonological categories. This can be observed in the exact timing of the effect ([a] at 95 ms, [u] at 120 ms; see Roberts et al., 2004), and in the location of the activity in the brain (Obleser et al., 2004).

The mismatch negativity (MMN) component of the ERP allows for some indirect insights into this early phase of processing. MMN is typically obtained in a classic passive oddball paradigm. In this experimental protocol, a sequence of identical sounds, the standards, (for example [a]), is interrupted occasionally by another sound, the deviant, (for example [u]), as in [a a a u ...]. A standard experimental design, called reversed oddball design, will test [a a a u ...] with standard [a] and deviant [u] as well as [u u u a ...] with standard [u] and deviant [a], both with a considerable

number of repetitions. The activities of standard [u], deviant [u], standard [a], and deviant [a] are then each averaged separately, and the difference waves are calculated from the ERPs either by subtracting the ERP of the original standard from the deviant ERP or by subtracting the ERPs elicited by the same stimulus when presented as standard and as deviant from the reversed oddball condition. A significant negative-going deflection in the difference wave calculated from the deviant and the standard ERP may be evidence for the MMN ERP component (Näätänen, 1992, 2001). This often occurs in the time range of 100–250 ms after the beginning of the deviating sound (e.g., Schröger, 2005).

Mismatch negativity studies also show the early effect of phonological categories (see for example Dehaene-Lambertz et al., 2000; Phillips et al., 2000); the evidence for this comes in part from comparisons between speakers of different languages (Näätänen et al., 1997; Winkler et al., 1999; Peltola et al., 2003). The speakers may react differently to a given sound contrast depending on the sound inventory of their native language.

### PREVIOUS MMN-STUDIES ON PHONOTACTIC RESTRICTIONS

Some other studies have investigated the effects of phonological rules or phonotactic constraints in MMN protocols. Dehaene-Lambertz et al. (2000) investigated the Japanese restriction that the syllable coda allows only place-assimilated nasals (and the first part of a geminate; cf. Itô, 1986). When Japanese listeners hear a sequence like [igmo] they perceive the presence of an additional vowel as in [igumo]. The additional vowel makes the sequence well-formed in Japanese. French speakers do not hear such an

additional vowel. In an MMN experiment, pairs like [igmo] and [igumo] were investigated for effects of the vowel epenthesis. In Japanese speakers, there was no MMN effect, while in French speakers there was. As the authors note, these results “suggest that the impact of phonotactics takes place early in speech processing and support models of speech perception, which postulate that the input signal is directly parsed into the native language phonological format” (p. 635). Since we were interested, in our own studies, in effects of processing that take place outside of the focus of attention, we mention that the participants of the study of Dehaene-Lambertz et al. (2000) were instructed to pay attention to the stimuli and to answer for each five-stimulus sequence whether the fifth (the deviant) was different from the preceding four.

Mitterer and Blomert (2003) investigated optional nasal place assimilation in Dutch compounds (in terms of lexical phonological theory: postlexical assimilation). They paired the unassimilated [tuinbank] with the assimilated [tuimbank] (both “garden bank”), both of which are possible forms of this word in Dutch, while participants were watching a silent movie. This contrast was compared with the pairing of [tuinstoel] and [tuimstoel]. While the difference between [tuin] and [tuim] was identical in the two stimulus pairs, the change was not motivated by assimilation in [tuimstoel]. A significant difference between standard and deviant was found in the latter pair, but not in the former pair where the assimilation process relates the two forms. Therefore, the regressive assimilation process is relevant to early pre-attentive processing.

Flagg et al. (2006) investigated an assimilatory nasalization process in English with an MEG study. In /ama/ the first vowel optionally gets nasalized by the following nasal as in [āma]. Flagg et al. (2006) classified this alternation as phonological assimilation, though we point out that the process more likely is to be seen as coarticulatory, i.e., phonetic in nature. Such a nasalized vowel was spliced before a non-nasal consonant as in [āba]. The participants of the experiment were watching silent movies during passive stimulation. A latency delay was found for the M50 response elicited by the incongruent plosive in [āba] compared to [aba]. This indicated that the nasalization process was relevant to very early pre-attentive processing.

Steinberg et al. (2010a,b, 2011) investigated a German allophonic alternation related to two dorsal fricative allophones both represented orthographically as “ch.” The palatal allophone of this fricative occurs after front vowels ([dɪçt] *dicht* ‘dense’) and the velar allophone after back vowels ([dɔxt] *Docht* ‘wick’). This alternation is also known as dorsal fricative assimilation (DFA). From a range of different experiments that all provide evidence for the effect of DFA in pre-attentive processing, we here choose one for presentation: the ill-formed non-word \*[ɛx] combines a velar fricative with a front vowel. Contrasted with the well-formed pseudoword [ɔx] as standard, there was a mismatch effect attributable to the different vowels. The fricatives are segmentally identical, so that the deviant [ɔx] did not show an MMN due to the fricative in the comparison condition. However, the ill-formed deviant \*[ɛx] elicited an additional MMN response attributable to the fricative. This response was temporally separated from the vowel-related

MMN and attributed to the abstract phonotactic ill-formedness of the deviant.

In the present study on final devoicing (FD) in German, we continue our investigation of bona fide productive lexical phonological rules in German, i.e., of alternations that apply obligatorily, without idiosyncratic exceptions and within the domain of words or pseudowords, but not across words or pseudowords.

## FINAL DEVOICING

Final devoicing operates on what has classically been analyzed as a voicing contrast (see e.g., Rubach, 1990; Hall, 1992). Jessen and Ringen (2002) have argued that the contrast instead involves the feature [spread glottis] for the plosives, and Beckman et al. (2009), building on this, have argued that the German fricatives are specified for both [spread glottis] and [voiced] (see also Vaux, 1998 for arguments that voiceless fricatives are specified [+spread glottis] across languages). In the present experiment, we employed a voicing distinction in fricatives. Assuming such a dual specification, we expect no effects of lexical underspecification, which have been argued to affect MMN by Eulitz and Lahiri (2004), Cornell et al. (2011, 2013), and Scharinger et al. (2012). Instead, voiced fricatives would be specified [+voiced] and voiceless ones would be specified [+spread glottis] in the mental lexical entries.

The German plosives and fricatives that allow such a laryngeal contrast, here transcribed in terms of voicing, are [p/b, t/d, k/g, f/v, s/z]. Both members of each pair can occur in the onset of a syllable before a vowel. In the classical analysis, the voiced values become voiceless in a syllable coda (Rubach, 1990; Hall, 1992). Thus, the two genitive forms [ra.d-əs] (*Rades* ‘wheel-GEN’) and [ra.t-əs] (*Rates* ‘advice-GEN’) distinguish [d] and [t] in the syllable onset before a vowel. However, in the nominative form, without the genitive suffix [-əs], the forms are identically pronounced [ra:t] (‘wheel’/‘advice’). Here, /d/ and /t/ are in the syllable coda and only the voiceless pronunciation [t] occurs. The change from /d/ in the mental lexical entry /rad/ to [t] in the pronunciation in coda position is called final devoicing (FD). There are different suggestions about the best way of describing and capturing the correct environment (see e.g., Lombardi, 1991, 1999; Steriade, 1997; Beckman et al., 2009). There is also a debate about whether the voicing neutralization is phonetically complete (see e.g., Port and O’Dell, 1985; Beckman et al., 2009). However, it is clear that the change takes place obligatorily in a set of core environments that include the word-final position, that there are no lexically marked exceptions, and that the change is bounded by the word.

Hwang et al. (2010) showed that English voicing agreement in consonant clusters as in their pseudoword stimuli [ʊts] and [ʊdz] lead to processing difficulties in non-agreeing clusters like \*[ʊds], which were not found in non-agreeing clusters like \*[ʊtz]. Poeppel and Monahan (2011, p. 947f.) refer to results of a related MEG experiment in which a distinction between [ʊts] and \*[ʊds] was found around 150 ms after the onset of the fricative. Hwang et al. (2010) interpret their results in terms of the underspecification for voicing of voiceless plosives in English postulated by Lombardi (1991, 1999): speakers predict a following voiced sound after [d] in \*[ʊds] but do not predict a following voiceless sound after [t] in \*[ʊtz] because [t] is underspecified for voicing. We think that



this explanation may apply to a phonological surface structure in which the position preceding the fricative is conceivably one of laryngeal neutralization (see Steriade, 1997). It is conceivable that the voicelessness of [t] preceding a fricative may be accounted for by laryngeal neutralization by the processing system, while the voicing of [d], if followed by a fricative, can only be licensed by agreement with the fricative. Our assumptions about the underlying featural specifications in German are thus not in conflict with these interesting results.

The voicing distinction between obstruents in German is phonetically implemented in several ways depending both on the manner of articulation and on the relative position of the sound. As we will focus on fricatives in intervocalic and final position in our study (see *Experimental design*), we will limit the following overview over phonetic voicing cues to these instances. Because in German, the phonetic implementations of the voicing contrast are – at least partly – neutralized in final positions, we also attend to voicing parameters obtained in languages like English in which FD is not operative. As shown by various phonetic studies (for an overview of the literature on the voicing distinction in German fricatives, see Jessen, 1998, pp. 65–66, 96; phonetic evidence on English fricatives is reviewed for instance by Stevens et al., 1992, and Maniwa and Jongman, 2009) the voicing distinction between fricatives is mainly coded by three kinds of parameters: first, the *duration of the fricative* (as reflected by the duration of friction noise in the acoustic signal) is shorter in voiced compared to voiceless fricatives. Preceding full vowels show – to some degree – the reversed durational pattern. Second, there are several spectral indicators for fricative voicing, most importantly the presence of periodic *low-frequency energy* during the fricative (reflecting vocal fold vibrations). Additionally, voiced fricatives are characterized by a lower Center of Gravity (COG) and higher variance compared to voiceless fricatives (cf. Maniwa and Jongman, 2009). Third, fricative voicing is indicated by a greater extent of the *F1 transitions* of preceding or following adjacent vowels (e.g., Stevens et al., 1992). Furthermore, vowels following a voiced fricative have been shown to begin with lower  $F_0$  than vowels after voiceless ones (Jessen, 1998).

### AN ASYMMETRY BETWEEN STANDARDS AND DEVIANTS

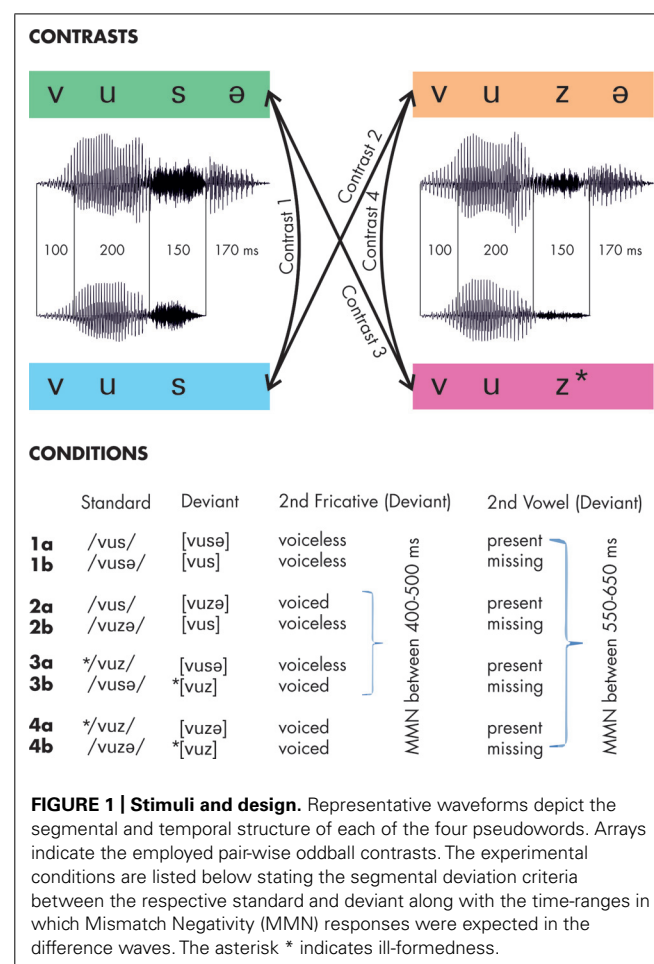
There is an interesting asymmetry in the roles played by standard and deviant in processing the oddball stimulation. Since the standard is repeated a number of times, and the pauses between the repetitions give sufficient time for it to be recognized as a particular phonological sound or sound sequence, it seems, put simply, that the expectation for another standard is phonologically represented, or represented in more abstract terms (Näätänen, 2001). The deviant, on the other hand, is just coming into the system and its initial processing is ongoing at the time when the mismatch against the standard arises.

Eulitz and Lahiri (2004), Cornell et al. (2011, 2013), and Scharinger et al. (2012) have argued that the standard can in certain ways be seen as similar to a mental lexical entry, likewise abstractly represented, and that the deviant can be seen as similar to the incoming acoustic information that the system seeks to match to an abstract lexical entry. They have argued that lexical underspecification of features matters for MMN in a way that can

be understood in these terms. A crucial aspect of the asymmetry for our experiment is that it provides a direction of application of FD: if it applies in an oddball protocol in early pre-attentive processing, it should apply in pairs in which the standard corresponds to a possible mental lexical entry to which FD could apply, and in which the deviant can be seen as similar to a spoken word to which FD has applied. For ease of exposition we therefore adopt some notation of Cornell et al. (2013). The standards are provided with slashes /./ and the deviants with squared brackets [.]. This is parallel to the phonological notation where /./ is used for mental lexical entries and [. ] for what is heard.

### EXPERIMENTAL DESIGN

The experiment reported here addressed German FD in pre-attentive phonological processing. The stimuli employed were [vus], \*[vuz], [vusə], and [vuzə] as depicted in **Figure 1**. We concentrated on four pair-wise contrasts each of which was employed twice with reversed roles of standard and deviant in the stimulations, resulting in a total of eight experimental conditions. As explained above, we marked the standard stimuli of the experimental conditions with /./ and the deviants with [.]. Our expectations were based on the similarities of standard stimuli to abstract phonological lexical representations on the one hand, and deviants to phonetic surface representations that are close to



the acoustic input on the other hand (Näätänen, 2001; Eulitz and Lahiri, 2004).

Contrast 1 is what would be an alternation in German for an underlying voiceless /s/. There is no change in voicing of the fricative. Hence, we expected MMN elicitation both for the additional vowel in 1a and for the missing vowel in 1b.

In contrast 2, stimuli differ with respect to the voicing of the second fricative. In condition 2a, this change is phonologically unmotivated. Furthermore, the deviant differs due to its additional vowel. Here, we expected an MMN to be elicited by each of these changes. Run the other way around, as in condition 2b, the whole contrast between standard and deviant could be interpreted as an alternation in German for an underlying voiced /z/ in /vuzə/, with FD in [vus]. While the phonetic differences were the same in 2a and 2b, we expected the respective MMN patterns to reflect that standard and deviant were phonologically related due to FD in 2b but not in 2a.

Contrasts 3 and 4 employ the ill-formed stimulus \*[vuz]. It is ill-formed because FD would obligatorily turn it into [vus] in German. In contrast 3, the deviant enters into an unmotivated voicing alternation of the second fricative in both conditions. In addition, the stimuli differed with respect to the presence or absence of the final vowel. Consequently, we expected to find MMN for each of these segmental differences in both conditions. Furthermore, we were interested in whether we would find an additional effect attributable to the ill-formedness of \*[vuz] when being presented as deviant. As the phonological violation \*[vuz] coincided with the absence of the second vowel, we expected these mismatch responses to overlay, as reflected by a larger MMN amplitude in condition 3b compared to 1b.

In contrast 4, there is no change in voicing with respect to the second fricative, so no effects were expected in any corresponding time window. With respect to the difference in the final vowel, we were interested in whether condition 4a would show reduced MMN compared to the remaining a-conditions; this may be expected as it would reflect a remedy of the violation of FD in the standard \*[vuz]. In condition 4b, we again expect superimposed mismatch effects due to the ill-formedness of the deviant \*[vuz] and due to the missing second vowel as in condition 3b.

## MATERIALS AND METHODS

### PARTICIPANTS

Sixteen volunteers participated in the study (four male; median age was 26 years, range from 22 to 33), all of them right-handed and monolingual native speakers of German. Handedness was assessed using an inventory adopted from Oldfield (1971). All participants reported normal auditory and normal or corrected-to-normal visual acuity and no neurological, psychiatric, or other medical problems. They gave informed written consent. The study conformed to The Code of Ethics of the World Medical Association (2013, Declaration of Helsinki).

### MATERIALS

As described, four pseudowords were used as stimuli: [vus], [vusə], \*[vuz], and [vuzə]. The stimuli are phonotactically well-formed in German, except for the non-word \*[vuz], which fails to have undergone FD. The stimuli [vus], [vusə], \*[vuz], and [vuzə] were

articulated numerous times by a professional male speaker with a fundamental frequency ( $F_0$ ) of about 100 Hz, and digitally recorded with a 48 kHz sampling rate and a 16 bit resolution using a RME Fireface 800 recording device (Audio AG, Haimhausen, Germany) and a Neumann U87 Microphone (Georg Neumann GmbH, Berlin, Germany).

Stimulus preparation for ERP-experiments on speech processing is always a compromise. The point is to control for lower-level acoustic stimulus characteristics in order to avoid confounds with higher-level linguistic factors while on the same time keeping the stimuli as natural as possible and avoiding artifacts caused by manipulation. To assure some acoustic variability of the stimulus material, we selected 5 different utterances of each pseudoword resulting in a set of 20 pseudoword stimuli in total (see Eulitz and Lahiri, 2004; Jacobsen et al., 2004; Steinberg et al., 2012). However, the conflicting methodological requirements mentioned above concern our study in a special way. The phonological issue under investigation (i.e., the voicing distinction between [s] and [z]) is also coded by the inherent durational differences both between the voiced and voiceless fricatives and between the preceding vowels. Other sufficient voicing cues for fricative perception are the presence or absence of low frequency energy during the fricative in the acoustic signal and distinct F1 transitions on vowel-fricative and fricative-vowel boundaries. These cues are also highly reliable at least in intervocalic and final fricative position (as in our stimuli).

Based on these considerations we decided to normalize the segmental durations of the stimuli across contrasts and to base the voicing distinction only on spectral phonetic parameters. Durational normalization was performed using the time-domain pitch synchronous overlap add (TD-PSOLA) algorithm provided by Praat software (Boersma and Weenink, 2010). Segmental durations were equated by setting the initial fricative to 100 ms (mean original durations of [v] in ms: [vus] 129, [vusə] 119, [vuz] 102, [vuzə] 109), the full vowel to 200 ms (mean original durations of [u] in ms: [vus] 200, [vusə] 188, [vuz] 285, [vuzə] 203), the second fricative to 150 ms (mean original durations of [s] in ms: [vus] 329, [vusə] 177; of [z] in ms: [vuz] 259, [vuzə] 119) and the final vowel to 170 ms (mean original durations of schwa in ms: [vusə] 189, [vuzə] 167). Afterward, intensities were normalized using the root mean square (RMS) of the whole sound file.

Theoretically, the duration normalization bore two risks: first, originally voiced fricatives might be perceived as voiceless after the relative lengthening of the fricative and the shortening the preceding vowel. Second, the contrary effect might have occurred to the originally voiceless fricatives. However, our ERP data clearly indicate that a distinction in the fricative has been detected in both directions in contrasts 2 and 3 [see *Analysis of the voicing change in the fricatives (contrasts 2 and 3)*]. Nevertheless, we performed acoustic analyses after the manipulation procedures to ensure that sufficient phonetic information was left in the stimulus material coding the voicing distinction between the fricatives [v] and [s] and to test potential interactions with the syllabic position of the fricative. We tested both offset F1 transitions of the first vowel, and the first two spectral moments of the fricative.

Formant measures were taken from each single stimulus file as mean values within 20 ms analysis windows by using the linear

prediction-based burg method (as implemented in Praat) with a pre-emphasis frequency of 50 Hz. F1 measures were taken from the mid part (190–210 ms) and from the final part of the vowel (280–300 ms) by automatically determining maximally two formants below 2000 Hz. F1-transitions were analyzed by means of a univariate mixed design analysis of variance (ANOVA) with the within-items factor TRANSITION (mid vowel/vowel offset) and the between-items factors FRICATIVE (voiceless/voiced) and SYLLABLE (mono-/bisyllabic). We found a main effect of the factor TRANSITION ( $F_{1,16} = 6.2$ ;  $p = 0.024$ ;  $\eta_p^2 = 0.279$ ) and a significant interaction TRANSITION\*FRICATIVE ( $F_{1,16} = 5.9$ ;  $p = 0.027$ ;  $\eta_p^2 = 0.269$ ). As expected from the literature, the first vowel formant showed a significantly falling pattern when preceding the voiced fricative (F1 mid vowel: 357 Hz/F1 vowel offset: 316 Hz; main effect TRANSITION in a broken down two-way ANOVA with TRANSITION and SYLLABLE:  $F_{1,8} = 13.8$ ;  $p = 0.006$ ;  $\eta_p^2 = 0.633$ ) while the F1 transition remained steady-state when being followed by the voiceless fricative (F1 mid vowel: 356 Hz/F1 vowel offset: 355 Hz; no significant effects). Note that there was no significant effect by the factor SYLLABLE.

To analyze the spectral qualities of the fricatives, FFT power spectra were calculated using a 50 ms Hann window that was centered over the mid part of the fricative (350–400 ms). From these spectra, COG and standard deviation (SD) were obtained. The spectral measures of the fricatives were analyzed by means of a multivariate ANOVA (MANOVA) with the between-items factors FRICATIVE and SYLLABLE as described before. A significant main effect of FRICATIVE indicates spectral differences between [s] and [z] (Pillai's trace = 0.532;  $F_{2,15} = 8.5$ ;  $p = 0.003$ ;  $\eta_p^2 = 0.532$ ). The factor SYLLABLE did not show any significant effects. The univariate analyses revealed that voiceless fricatives were characterized by significantly higher COG frequencies ([s] 7712 Hz/[z] 5908 Hz:  $F_{1,16} = 10.9$ ;  $p = 0.004$ ;  $\eta_p^2 = 0.406$ ), and lower SD ([s] 2094 Hz/[z] 2730 Hz:  $F_{1,16} = 14.0$ ;  $p = 0.002$ ;  $\eta_p^2 = 0.466$ ) compared to the voiced fricatives. Based on this we assumed that the voicing distinction in the stimulus material was sufficiently coded phonetically even though durational voicing cues had been neutralized by manipulation.

## EXPERIMENTAL DESIGN AND PROCEDURE

As described above, four different experimental contrasts were employed: [vus] vs. [vusə] (contrast 1), [vus] vs. [vuzə] (contrast 2), \*[vuz] vs. [vusə] (contrast 3) and \*[vuz] vs. [vuzə] (contrast 4). Each pair-wise contrast was presented twice in oddball sequences, both using one pseudoword as standard (85% of the trials = 1360 items) and the other as deviant and the other way around (reversed oddball-design), resulting in eight experimental conditions. Oddball sequences of 1600 trials in total were presented per condition, using all tokens of each pseudoword equally. Standard and deviant stimuli were delivered in pseudo-randomized order forcing at least two standards to be presented between successive deviants. Oddball conditions were then divided into two technical blocks each, resulting in a total of 16 stimulation blocks per participant. Sessions were split into two parts, so the second half of each condition was presented on a second day. Stimulus sequences were presented with a stimulus onset

asynchrony randomly varying from 550 to 900 ms in units of 10 ms. The order of the experimental blocks was counterbalanced between participants. Participants were seated comfortably in a sound-attenuated and electrically shielded experimental chamber, and they were instructed to ignore the auditory stimulation while watching a self-selected silent subtitled movie. Stimuli were presented binaurally at 53 dB SPL through headphones (Sennheiser HD 25-1 II; Sennheiser electronic GmbH & Co. KG, Wedemark, Germany). Loudness was measured by means of an artificial head (artificial head HMS III.2; HEAD acoustics GmbH, Herzogenrath, Germany). All participants reported that they were able to ignore the auditory stimulation. Informal questioning of the participants revealed that they had perceived all stimulus types as speech sounds. A whole experimental session lasted approximately 180 min (plus additional time for electrode application and removal) including ten short breaks of about 2 min each.

## ELECTROPHYSIOLOGICAL RECORDINGS

The EEG (Ag/AgCl electrodes, Falk Minow Services, V-Amp EEG amplifier; Brain Products GmbH, Gilching, Germany) was recorded continuously from 26 standard scalp locations according to the extended 10–20 system (American Encephalographic Society, 1994; FP1, FP2, F7, F3, FZ, F4, F8, FC5, FC1, FCZ, FC2, FC6, C3, CZ, C4, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2) and from the left and right mastoids. The reference electrode was placed on the tip of the nose, and an additional electrode placed at AFZ was used as ground during recording. Electroocular activity was recorded with two bipolar electrode pairs, the vertical electrooculogram (EOG) was obtained from the right eye by one supraorbital and one infraorbital electrode and the horizontal EOG from electrodes placed lateral to the outer canthi of both eyes. Impedances were kept below 10 k $\Omega$ . On-line band-pass filtering of the EEG and EOG signals was carried out using a 0.011 Hz high-pass and a 100 Hz low-pass filter. The signal was digitized with a 16 bit resolution at a sampling rate of 500 Hz.

## DATA ANALYSIS

Off-line signal processing was carried out using EEP 3.0 (ANT Neuro, Enschede, Netherlands). EEG-data were band-pass filtered with a finite impulse response filter: 4001 points, critical frequencies of 0.5 Hz (high-pass) and 15 Hz (low-pass; cf. Schröger, 2005). EEG epochs with a total length of 1050 ms, time-locked to the onset of the stimuli and including a 100 ms pre-stimulus baseline, were extracted and averaged separately for each stimulus probability (standard, deviant), for each pseudoword, and for each participant.

The ERP responses to the first five stimuli per block as well as to each standard stimulus immediately following a deviant were not included in the analysis. Epochs showing an amplitude change exceeding 100  $\mu$ V at any of the recording channels were rejected. In the present study, an average of 15.1% (SD 6.2%) of the trials per participant was rejected prior to ERP computation. Grand-averages were subsequently computed from the individual-subject averages.

To quantify the full MMN amplitude, the scalp ERPs were re-referenced to the averaged signal recorded from the electrodes



positioned over the left and right mastoids. This computation results in an integrated measure of the total neural activity underlying the auditory MMN (e.g., Schröger, 2005).

Deviant-minus-standard difference waveforms were calculated for each pseudoword per oddball condition by subtracting the ERPs elicited by the standard point by point from the ERPs elicited by the original deviant obtained from the same oddball condition, i.e., the MMN elicited by [vusə] in condition 1a was quantified as difference between the deviant ERP from [vusə] and the standard ERP from [vus]. We opted for original contrasts from the same block in order to prevent superimposing effects from the block context to affect our comparisons.

Deviance-related effects (as the MMN) were quantified by measuring the ERP amplitudes as mean voltages in a fixed analysis window of 40 ms (for the width of the analysis window, cf. Luck, 2005, pp. 234). These windows were adjusted a posteriori on the basis of the grand-averaged deviance-minus-standard difference waves (cf. Picton et al., 2000). We adjusted separate windows for each condition and for each deviation by identifying the peak latencies of any distinguishable negative-going deflection (averaged across F3, Fz, F4, C3, Cz, and C4 electrode positions) within *a priori* determined time ranges. First, any effect due to the voicing alternation in the second fricative was expected to occur between 400 and 500 ms post stimulus onset (note that this latency equals 100 to 200 ms after the onset of the differing fricatives). This voicing alternation only occurs in contrasts 2 and 3. Second, deviations due to the presence or absence of the second vowel were expected to affect processing within the time range of 550–650 ms post stimulus onset (i.e., 100–200 ms after the offset of the fricative/onset of the final vowel). In singular cases, additional earlier or later time windows were analyzed in an exploratory approach.

Statistical analyses were performed with SPSS (IBM SPSS Statistics 21). As the MMN is known to be maximal over frontal scalp areas (cf. Kujala et al., 2007), we decided to base our analyses only on the F-line positions by collapsing the ERPs obtained at F3, Fz, and F4 into one single measure. Separately for each analysis window, an overall univariate repeated-measures analysis of variance, henceforth ANOVA, was run including the within-subjects factors STIMULUS PROBABILITY (standard/deviant), CONTRAST (depends on the window), and VOWEL (additional vowel in the deviant is present/missing). Afterward, analyses were broken down if appropriate. Finally, comparisons between conditions relating to the hypotheses were performed using repeated-measures ANOVAs with the factors introduced above. Only significant main effects of the factor STIMULUS PROBABILITY and interactions with this factor were reported. The level of type 1 error was set to  $p < 0.05$  and, in case of multiple *post hoc* comparisons, Bonferroni correction was applied. If the sphericity assumption was violated (indicated by the Mauchly test), the original degrees of freedom were provided along with the Greenhouse-Geisser-epsilon. Finally, partial eta-squared ( $\eta_p^2$ ) effect sizes were given for all significant effects.

## RESULTS

The ERP results for all conditions are depicted in **Figure 2**. Also, this figure shows the respective analysis windows for each

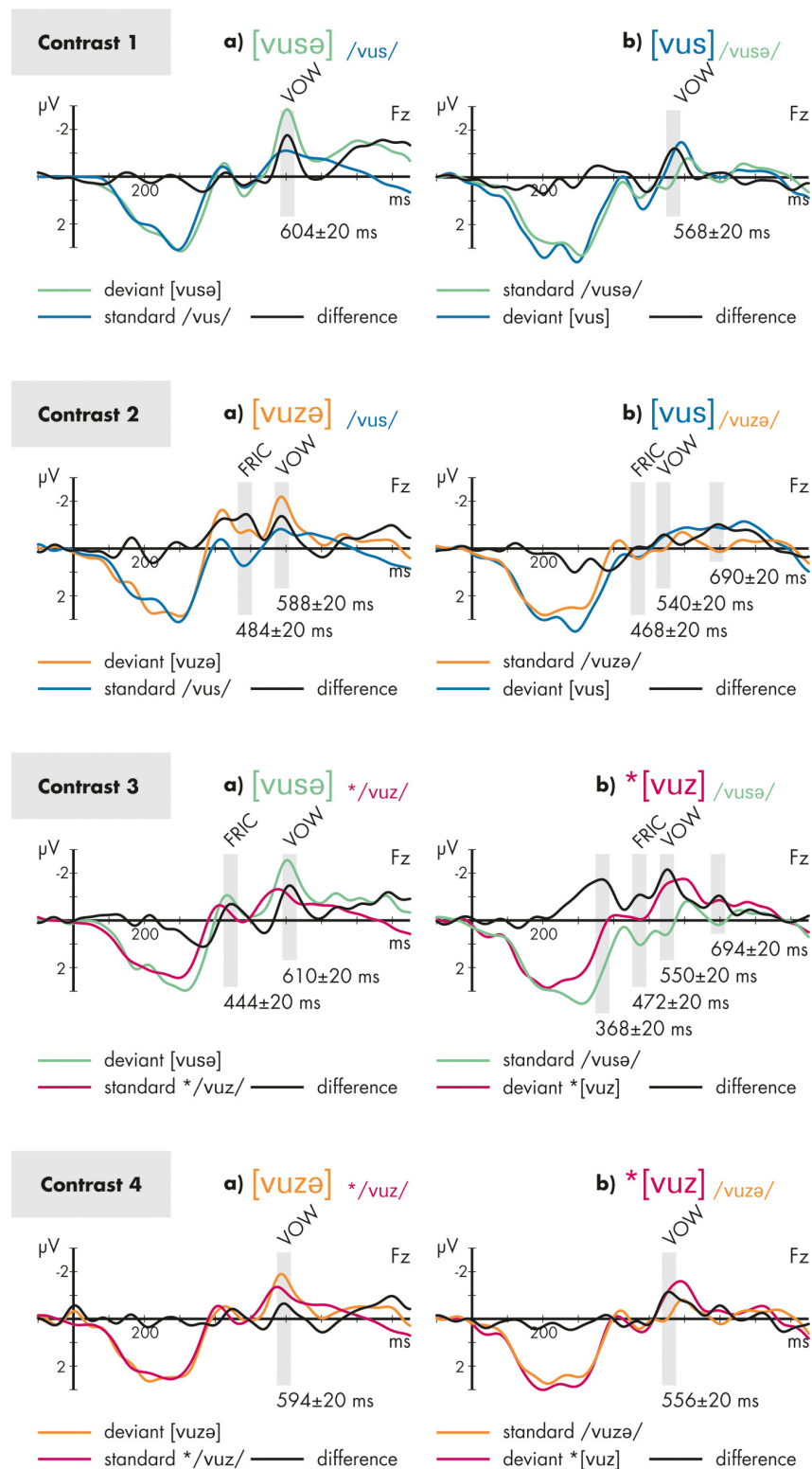
effect. The outcomes of the statistical analyses based on these windows are presented below separately for each analysis window. In **Figure 3**, topographical maps of the analyzed MMN effects are provided separately for each condition and time window.

### ANALYSIS OF THE VOICING CHANGE IN THE FRICATIVES (CONTRASTS 2 AND 3)

For the MMN responses to the fricatives (FRIC in **Figure 2**) the overall ANOVA with the factors STIMULUS PROBABILITY (standard/deviant), VOWEL (additional/missing), and CONTRAST (2/3) revealed a significant main effect of the factor STIMULUS PROBABILITY ( $F_{1,15} = 17.9$ ;  $p = 0.001$ ;  $\eta_p^2 = 0.544$ ), indicating the presence of an MMN across all conditions, and a significant interaction STIMULUS PROBABILITY\*VOWEL\*CONTRAST ( $F_{1,15} = 5.9$ ;  $p = 0.028$ ;  $\eta_p^2 = 0.284$ ), indicating different amplitudes of the MMN responses across conditions. Broken-down analyses were calculated separately for each contrast: in contrast 2, the main effect for STIMULUS PROBABILITY ( $F_{1,15} = 9.1$ ;  $p = 0.009$ ;  $\eta_p^2 = 0.387$ ), and also the interaction STIMULUS PROBABILITY\*VOWEL ( $F_{1,15} = 5.9$ ;  $p = 0.028$ ;  $\eta_p^2 = 0.282$ ) were significant, the latter indicating a stronger MMN response in condition 2a compared to condition 2b. In contrast 3, only a significant main effect for STIMULUS PROBABILITY was obtained ( $F_{1,15} = 9.3$ ;  $p = 0.008$ ;  $\eta_p^2 = 0.384$ ).

### ANALYSIS OF THE EFFECT DUE TO THE CHANGE IN THE FINAL VOWEL (ALL CONTRASTS)

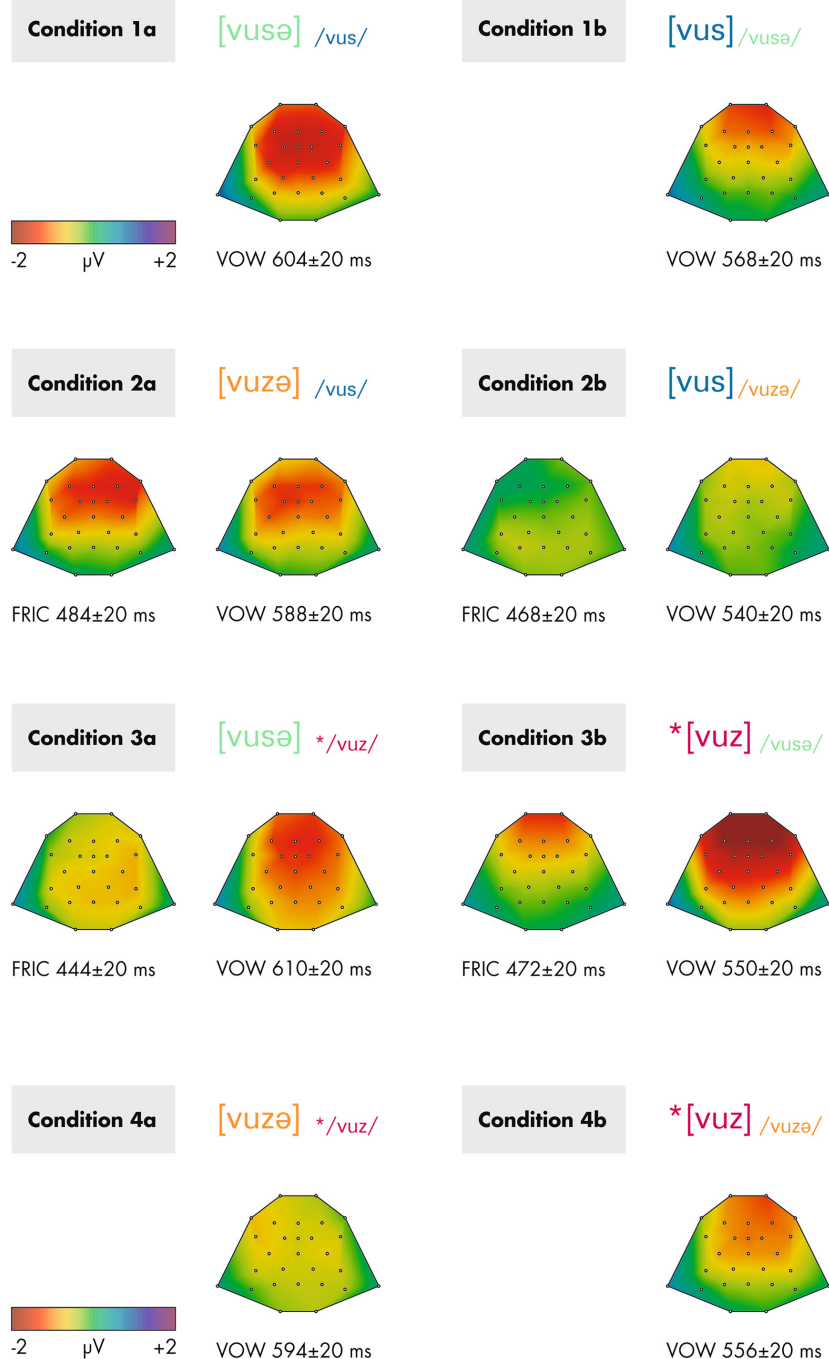
For the MMN responses to the additional or missing vowel (VOW in **Figure 2**) the overall ANOVA with the factors STIMULUS PROBABILITY (standard/deviant), VOWEL (additional/missing), and CONTRAST (1/2/3/4) revealed a significant main effect for STIMULUS PROBABILITY ( $F_{1,15} = 25.0$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.625$ ), as well as significant interactions STIMULUS PROBABILITY\*CONTRAST ( $F_{3,45} = 4.0$ ;  $p = 0.026$ ;  $\epsilon = 0.725$ ;  $\eta_p^2 = 0.209$ ) and STIMULUS PROBABILITY\*CONTRAST\*VOWEL ( $F_{3,45} = 3.1$ ;  $p = 0.039$ ;  $\epsilon = 0.928$ ;  $\eta_p^2 = 0.172$ ). Next, analyses were broken down by the factor VOWEL. Comparing the MMN amplitudes for the a-conditions, only a significant main effect for STIMULUS PROBABILITY ( $F_{1,15} = 9.3$ ;  $p = 0.008$ ;  $\eta_p^2 = 0.382$ ) was found, but no interaction with this factor. For the b-conditions, the main effect STIMULUS PROBABILITY ( $F_{1,15} = 21.1$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.585$ ) and the interaction STIMULUS PROBABILITY\*CONTRAST ( $F_{3,45} = 6.2$ ;  $p = 0.002$ ;  $\epsilon = 0.880$ ;  $\eta_p^2 = 0.292$ ) were significant. This interaction indicates differences in MMN amplitudes due to the missing final vowel across the contrasts. We *a priori* were only interested in potential differences between conditions 1b and 3b, both sharing the same legal standard /vusə/. A broken-down ANOVA with STIMULUS PROBABILITY and CONTRAST (1/3) revealed a significant main effect for STIMULUS PROBABILITY ( $F_{1,15} = 35.4$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.703$ ), and a significant interaction between both factors ( $F_{1,15} = 5.1$ ;  $p = 0.039$ ;  $\eta_p^2 = 0.252$ ), indicating stronger MMN amplitudes for 3b compared to 1b.



**FIGURE 2 | Grand-averaged and re-referenced event-related potentials (ERPs) elicited by the stimuli of Contrasts 1–4 depicted separately for conditions a (left) and b (right) at Fz electrode site.** The color of the ERPs codes the stimulus that elicited that ERP. The black line represents the difference wave calculated for each

condition. The gray bars indicate the time windows of statistical analyses. FRIC indicates that the marked time range is attributed to the voicing change in the second fricative, VOW indicates that the marked time range is attributed to the final vowel. The asterisk \* indicates ill-formedness.





**FIGURE 3 |**Topographical maps of the analyzed deviance-minus-standard differences obtained from grand-averaged re-referenced data within the time windows stated below each map. Depicted are contrasts 1–4 separately for conditions a (left) and b (right). FRIC

indicates that the marked time range is attributed to the voicing change in the second fricative, VOW indicates that the marked time range is attributed to the final vowel. The asterisk \* indicates ill-formedness.

### EXPLORATIVE ANALYSES OF EARLIER AND LATER EFFECTS (CONTRASTS 2 AND 3)

In conditions 2b and 3b, unexpected deviance-related effects were found in a time range later than 650 ms post stimulus onset. These effects were analyzed as described above: a significant main

effect of STIMULUS PROBABILITY ( $F_{1,15} = 19.5$ ;  $p = 0.001$ ;  $\eta_p^2 = 0.565$ ) was found but no interactions with this factor. Because of its latency, it seems possible to us that this effect reflects morphological processing (see Royle et al., 2010). This is conceivable if the additional vowel is processed as a morphological suffix.

Furthermore, a strong deviance-related effect was observed in condition 3b that appeared in an unexpected early time range before 400 ms, i.e., before the onset of the deviating fricative. Because of its latency, this effect seemed to be temporally related to the later part of the first vowel [u]. This effect was compared with a corresponding time window in condition 1b ( $360 \pm 20$  ms) that shared the legal standard stimulus /vusə/. A significant main effect STIMULUS PROBABILITY ( $F_{1,15} = 39.1$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.722$ ) was found as well as a significant interaction STIMULUS PROBABILITY\*CONTRAST ( $F_{1,15} = 9.0$ ;  $p = 0.009$ ;  $\eta_p^2 = 0.374$ ), indicating a stronger deviance-related response in 3b compared to 1b.

## DISCUSSION

### REMARKS ON CONTRAST 1

Our statistical assessment employed condition 1b as a comparison condition for condition 3b (see *Evidence for the relevance of final devoicing in condition 3*). However, contrast 1 (pairing the legal stimuli [vusə] and [vus] with no voicing change) is here also briefly considered on its own. Visual inspection of the difference waves for contrast 1 in **Figure 2** shows distinct MMN responses that are attributable to the presence vs. absence of the final vowel, but no further effects, in particular no effects between 450 and 550 ms where differences attributable to the fricative would occur. This provides some assurance that effects attributed to fricatives in other conditions were not general consequences of our stimulus contrasts in which stimuli with and without a final vowel are compared. There is, for example, a distinction in syllabification. The a-conditions are syllabified like 1a: [vu.se] while the b-conditions are single syllables like 1b: [vus]. This distinction could in principle have phonetic correlates in regard to the extent of coarticulation of the [s/z] with the preceding vowel. Recall that the phonetic analysis of the stimuli did not detect any such differences. Condition 1 suggests that such differences, if they exist after all, also did not lead to observable effects in the difference wave.

### EVIDENCE FOR THE RELEVANCE OF FINAL DEVOICING IN CONTRAST 2

The following sketch shows condition 2b next to condition 2a. We included a dot to mark the syllable boundary in [vu.zə].

	Standard	Deviant	Comment
2a:	/vus/	[vu.zə]	significantly stronger effect for voicing change than in 2b
2b:	/vu.zə/	[vus]	related by final devoicing

There is a significant difference between conditions 2a and 2b in the processing correlates of the voicing change in the fricative. The MMN effect due to the voicing mismatch in condition 2a was absent in condition 2b, where the voicing change was motivated by FD. This significant difference between conditions 2a and 2b is here interpreted as evidence for the relevance of FD in pre-attentive processing.

### REMARKS ABOUT REACTIONS TO THE FINAL VOWEL IN CONDITION 2b

We turn to some remarks about the MMN response due to the additional/missing vowel in conditions 2a and 2b. The plots in

**Figure 2** suggest that the response attributable to the missing final vowel in the deviant of condition 2b was also reduced. We here want to comment this impression for the benefit of possible future experiments that might investigate such an effect more specifically. The observation suggests that the expectation of any upcoming auditory event, which is violated in the deviant and shown by the MMN, is not limited to the expectation of just another standard stimulus. It seems, instead, that this expectation can be modulated by what is found earlier in the deviant. The system seems to have related /vuzə/ and [vus] by FD. If the system possesses knowledge of the environment of FD, it will then expect the absence of a vowel following [vus], since FD would not have applied in the presence of a following vowel. (Similar expectations could also be modulated by phonetic factors that might allow the anticipation of the absence of a final vowel. However, the reduced MMN response to the missing final vowel seems to be specific to condition 2b, where FD has applied.) It is also possible, then, that the standard /vuz+ə/ and the deviant [vus] were processed as morphologically related by the omission of an inflectional element [ə] in the deviant, with phonological adjustment due to FD. It seems conceivable that this was related to the late deviance-related effect that was observed about 250 ms after the missing vowel had become detectable.

We note that we have argued (Jacobsen et al., 2013) against the assumption of successive MMN responses in case of mismatching monosyllabic vowel-consonant sequences, where both the vowel and the consonant differed. However, the case at hand is different in an important aspect: the second deviation in the present contrast pairs, namely the missing or additional final vowel in contrasts 2 and 3, did not just involve a distinct sound, but established a distinction in syllable structure between standard and deviant. By this, the present stimulus contrasts were clearly different not just at the segmental but also at suprasegmental representation levels.

### EVIDENCE FOR THE RELEVANCE OF FINAL DEVOICING IN CONDITION 3

It was seen in the presentation of the results that condition 3b and condition 1b both have MMN responses attributable to the missing vowel, and that both effects furthermore differ significantly in strength. This is illustrated in the following sketch.

	Standard	Deviant	Comment
1b:	/vusə/	[vus]	significant effect for missing final vowel
3b:	/vusə/	*[vuz]	significantly stronger effect for missing final vowel

It was suggested that this is evidence for a superposed effect of the ill-formedness of the deviant \*[vuz] in condition 3b, which becomes manifest in the signal simultaneously with the absence of the final vowel. This distinction provides further evidence for the relevance of FD in pre-attentive processing.

### REMARKS ON REACTIONS TO THE FIRST VOWEL IN CONDITION 3

The comparison between conditions 1b and 3b is repeated in the following, this time highlighting a significant distinction that was found post hoc: condition 3b showed an effect at the time at which

the second part of the vowel [u] is expected to be processed. The distinction to 1b was seen to be significant.

	Standard	Deviant	Comment
1b:	/vusə/	[vus]	no processing effect attributable to [u]
3b:	/vusə/	*[vuz]	significant effect attributable to [u]

This effect in 3b may be related to the anticipation of [z] during the vowel [u] due to coarticulatory cues. It is furthermore possible that phonetic factors allowed an early prediction of the syllable structure. The system might have noticed in \*[vuz] during the vowel that there would be an upcoming voiced fricative within the same syllable, in violation of FD. If so, the early strong MMN effect before 400 ms in condition 3b might already be a first electrophysiological response to the ill-formedness of the deviant.

## SUMMARY

In summary, we have found two pieces of evidence for the role of FD in pre-attentive processing. While condition 2a, [vuza]/vus/, showed mismatch effects due to the voicing change in the fricative, these are significantly reduced (and in fact absent) in condition 2b, [vus]/vuza/, in which the two forms can be related by FD. In condition 3b, \*[vuz]/vusə/, an overlaid effect of the violation of FD in the deviant \*[vuz] was found.

An interesting aspect of our findings is that they provide evidence that syllable-related lexical phenomena such as FD are already taken into account by the processing system in an early pre-attentive stage. This point is new insofar the only previous study we are aware of that showed the processing relevance of a syllable-related process is Dehaene-Lambertz et al. (2000), which did not employ a pre-attentive protocol.

## ACKNOWLEDGMENTS

This work was supported by the DFG SPP 1234 grant JA1009/10-2 to Thomas Jacobsen and Hubert Truckenbrodt and by the German Federal Ministry of Education and Research (BMBF), Grant Nr. 01UG1411. The authors are grateful to Jana Burock, Susan Beudt, Johannes Frey, Svantje Kähler, Aquiles Luna-Rodriguez, Jonathan Manske, Falco Walther, Mike Wendt and Lena Zielonka for technical help and valuable comments.

## REFERENCES

- American Encephalographic Society. (1994). Guideline thirteen: guidelines for standard electrode position nomenclature. *J. Clin. Neurophysiol.* 11, 111–113. doi: 10.1097/00004691-199401000-00014
- Beckman, J., Jessen, M., and Ringen, C. (2009). German fricatives: coda devoicing or positional faithfulness? *Phonology* 26, 231–268. doi: 10.1017/S0952675709990121
- Boersma, P., and Weenink, D. (2010). *Praat: Doing Phonetics by Computer (Version 5.1.44) [Computer Program]*. Available at: <http://www.praat.org/> [accessed October 20, 2010]
- Cornell, S. A., Lahiri, A., and Eulitz, C. (2011). What you encode is not necessarily what you store: evidence for sparse feature representations from mismatch negativity. *Brain Res.* 1394, 79–89. doi: 10.1016/j.brainres.2011.04.001
- Cornell, S. A., Lahiri, A., and Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: evidence from mismatch negativity. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 757–772. doi: 10.1037/a0030862
- Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *J. Cogn. Neurosci.* 12, 635–647. doi: 10.1162/089892900562390
- Eulitz, C., and Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *J. Cogn. Neurosci.* 16, 577–583. doi: 10.1162/089892904323057308
- Flagg, E. J., Cardy, J. E. O., and Roberts, T. P. L. (2006). MEG detects neural consequences of anomalous nasalization in vowel-consonant pairs. *Neurosci. Lett.* 397, 263–268. doi: 10.1016/j.neulet.2005.12.034
- Hall, T. A. (1992). *Syllable Structure and Syllable-Related Processes in German*. Tübingen: Max Niemeyer. doi: 10.1515/9783111657882
- Hwang, S.-O. K., Monahan, P., and Isard, W. (2010). Underspecification and asymmetries in voicing perception. *Phonology* 22, 205–224. doi: 10.1017/S0952675710000102
- Itô, J. (1986). *Syllable Theory in Prosodic Phonology*. New York: Garland Publishing.
- Jacobsen, T., Schröger, E., and Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology* 41, 654–659. doi: 10.1111/1469-8986.2004.00175.x
- Jacobsen, T. K., Steinberg, J., Truckenbrodt, H., and Jacobsen, T. (2013). Mismatch negativity (MMN) to successive deviants within one hierarchically structured auditory object. *Int. J. Psychophysiol.* 87, 1–7. doi: 10.1016/j.ijpsycho.2012.09.012
- Jessen, M. (1998). *Phonetics and Phonology of Tense and Lax Obstruents*. Amsterdam: John Benjamins.
- Jessen, M., and Ringen, C. (2002). Laryngeal features in German. *Phonology* 19, 189–218. doi: 10.1017/S0952675702004311
- Kujala, T., Tervaniemi, M., and Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biol. Psychol.* 74, 1–19. doi: 10.1016/j.biopsycho.2006.06.001
- Lombardi, L. (1991). *Laryngeal Features and Laryngeal Neutralization*. Ph.D. thesis, University of Massachusetts, Amherst.
- Lombardi, L. (1999). Positional faithfulness and voicing assimilation in optimality theory. *Nat. Lang. Linguist. Theory* 17, 267–302. doi: 10.1023/A:1006182130229
- Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.
- Maniwa, K., and Jongman, A. (2009). Acoustic characteristics of clearly spoken English fricatives. *J. Acoust. Soc. Am.* 125, 3962–3973. doi: 10.1121/1.2990715
- Mitterer, H., and Blomert, L. (2003). Coping with phonological assimilation in speech perception: evidence for early compensation. *Percept. Psychophys.* 65, 956–969. doi: 10.3758/BF03194826
- Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale: Lawrence Erlbaum.
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38, 1–21. doi: 10.1111/1469-8986.3810001
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huottilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434. doi: 10.1038/385432a0
- Obleser, J., Lahiri, A., and Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J. Cogn. Neurosci.* 16, 31–39. doi: 10.1162/089892904322755539
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Peltola, M. S., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., and Näätänen, R. (2003). Native and foreign vowel discrimination as indexed by the mismatch negativity (MMN) response. *Neurosci. Lett.* 352, 25–28. doi: 10.1016/j.neulet.2003.08.013
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *J. Cogn. Neurosci.* 12, 1038–1055. doi: 10.1162/08989290051137567
- Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R. Jr., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37, 127–152. doi: 10.1111/1469-8986.3720127
- Poeppel, D., and Monahan, P. J. (2011). Feedforward and feedback in speech perception: revisiting analysis by synthesis. *Lang. Cogn. Process.* 26, 935–951. doi: 10.1080/01690965.2010.493301

- Port, R. F., and O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *J. Phon.* 13, 455–471.
- Roberts, T. P., Flagg, E. J., and Gage, N. M. (2004). Vowel categorization induces departure of M100 latency from acoustic predictions. *Neuroreport* 15, 1679–1682. doi: 10.1097/01.wnr.0000134928.96937.10
- Royle, P., Drury, J. E., Bourguignon, N., and Steinhauer, K. (2010). “Morphology and word recognition: an ERP approach,” in *Proceedings of the 2010 Annual Conference of the Canadian Linguistic Association*, Concordia University, Montreal. Available at: <http://homes.chass.utoronto.ca/~cla-acl/actes2010/actes2010.html>
- Rubach, J. (1990). Final devoicing and cyclic syllabification in German. *Linguist. Inq.* 21, 79–94.
- Scharinger, M., Bendixen, A., Trujillo-Baretto, N. J., and Obleser, J. (2012). A sparse neural code for some speech sounds but not for others. *PLoS ONE* 7:e40953. doi: 10.1371/journal.pone.0040953
- Schröger, E. (2005). The mismatch negativity as a tool to study auditory processing. *Acta Acust. United Acust.* 91, 490–501.
- Sharma, A., and Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *J. Acoust. Soc. Am.* 107, 2697–2703. doi: 10.1121/1.428655
- Steinberg, J., Truckenbrodt, H., and Jacobsen, T. (2010a). Preattentive phonotactic processing as indexed by the mismatch negativity. *J. Cogn. Neurosci.* 22, 2174–2185. doi: 10.1162/jocn.2009.21408
- Steinberg, J., Truckenbrodt, H., and Jacobsen, T. (2010b). Activation and application of an obligatory phonotactic constraint in German during automatic speech processing is revealed by human event-related potentials. *Int. J. Psychophysiol.* 77, 13–20. doi: 10.1016/j.ijpsycho.2010.03.011
- Steinberg, J., Truckenbrodt, H., and Jacobsen, T. (2011). Phonotactic constraint violations in German grammar are detected automatically in auditory speech processing: a human event-related potentials study. *Psychophysiology* 48, 1208–1216. doi: 10.1111/j.1469-8986.2011.01200.x
- Steinberg, J., Truckenbrodt, H., and Jacobsen, T. (2012). The role of stimulus cross-splicing in an event-related potentials study. Misleading formant transitions hinder automatic phonological processing. *J. Acoust. Soc. Am.* 131, 3120–3140. doi: 10.1121/1.3688515
- Steriade, D. (1997). *Phonetics in Phonology: The Case of Laryngeal Neutralization*. Manuscript, UCLA. Available at: <http://www.linguistics.ucla.edu/people/steriade/papers/PhoneticsInPhonology.pdf>
- Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., and Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *J. Acoust. Soc. Am.* 91, 2979–3000. doi: 10.1121/1.402933
- Vaux, B. (1998). The laryngeal specifications of fricatives. *Linguist. Inq.* 29, 497–511. doi: 10.1162/002438998553833
- Winkler, I., Lehtokoski, A., Alko, P., Vainio, M., Czigler, I., Csépe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cogn. Brain Res.* 7, 357–369. doi: 10.1016/S0926-6410(98)00039-1
- World Medical Association. (2013). *World Medical Association Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects*. Available at: <http://www.wma.net/en/30publications/10policies/b3/>

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 May 2014; accepted: 29 October 2014; published online: 25 November 2014.

Citation: Truckenbrodt H, Steinberg J, Jacobsen TK and Jacobsen T (2014) Evidence for the role of German final devoicing in pre-attentive speech processing: a mismatch negativity study. *Front. Psychol.* 5:1317. doi: 10.3389/fpsyg.2014.01317

This article was submitted to Language Sciences, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Truckenbrodt, Steinberg, Jacobsen and Jacobsen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Do syllables play a role in German speech perception? Behavioral and electrophysiological data from primed lexical decision

Heidrun Bien<sup>1</sup>, Jens Bölte<sup>2</sup> and Pienie Zwitserlood<sup>2\*</sup>

<sup>1</sup> Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK

<sup>2</sup> Institute for Psychology, Westfälische Wilhelms-University Münster, Münster, Germany

## Edited by:

Ulrike Domahs, University of Marburg, Germany

## Reviewed by:

Ulrike Schild, University of Tübingen, Germany

Markus Conrad, Freie Universität Berlin, Germany

## \*Correspondence:

Pienie Zwitserlood, Institute for Psychology, Westfälische Wilhelms-University Münster, Fliednerstr. 21, 48149 Münster, Germany  
e-mail: zwitser@uni-muenster.de

We investigated the role of the syllable during speech processing in German, in an auditory-auditory fragment priming study with lexical decision and simultaneous EEG registration. Spoken fragment primes either shared segments (related) with the spoken targets or not (unrelated), and this segmental overlap either corresponded to the first syllable of the target (e.g., /teis/ – /teisti/), or not (e.g., /teis/ – /teistləs/). Similar prime conditions applied for word and pseudoword targets. Lexical decision latencies revealed facilitation due to related fragments that corresponded to the first syllable of the target (/teis/ – /teisti/). Despite segmental overlap, there were no positive effects for related fragments that mismatched the first syllable. No facilitation was observed for pseudowords. The EEG analyses showed a consistent effect of relatedness, independent of syllabic match, from 200 to 500 ms, including the P350 and N400 windows. Moreover, this held for words and pseudowords that differed however in the N400 window. The only specific effect of syllabic match for related prime–target pairs was observed in the time window from 200 to 300 ms. We discuss the nature and potential origin of these effects, and their relevance for speech processing and lexical access.

**Keywords:** speech perception, form priming, ERPs, lexical access, lexical decision, syllables, fragment priming, German language

## INTRODUCTION

In a familiar language, listeners perceive speech as a sequence of discrete and meaningful units, though the spoken input consists of a continuous and often noisy signal. Speakers provide few reliable cues on how to organize this continuous signal into units of meaning. Speech is highly variable between, and even within, speakers. Moreover, speech segments (such as phonemes) partially overlap due to coarticulation, and can vary widely depending on the phonemic context. A question that is still not fully resolved is how this variable and noisy input is mapped onto word forms and meaning. One idea is that the input is mapped onto stored sublexical units, which aid access to lexical representations of word form. Among the candidates proposed as mediators between the acoustic input and the lexicon, two have received special attention: phonemes and syllables (Cutler et al., 1986; Dumay et al., 2002; Zwitserlood, 2004).

There is quite some evidence for phoneme-sized prelexical representations (cf. Hickok and Poeppel, 2007; Obleser and Eisner, 2009), including our own work with MEG and EEG (Bien et al., 2009; Bien and Zwitserlood, 2013). Obviously, sublexical phonemic units aid in abstracting away from the noisy input. According to word-recognition models in which the speech input, in terms of features and/or phonemic segments, is continuously mapped onto lexical word-form representations, irrespective of where words begin or end, such units suffice for lexical access and selection (cf. McClelland and Elman, 1986; see Christiansen and

Chater, 2001). Other models of spoken-word recognition assume that potential word onsets are important for lexical access (cf. Marslen-Wilson, 1987; Norris and McQueen, 2008). Detection of potential word-onsets presupposes segmentation or marking of incoming speech. Various cues have been proposed to signal potential onsets, such as transitional probabilities between consecutive segments (Saffran et al., 1996), phrase-boundary cues (Christophe et al., 2004), syllable duration and stress (Tyler and Cutler, 2009; Langus et al., 2012) or a combination of such cues (Mattys et al., 2005). Syllabic boundaries may also provide valuable cues for lexical access (Content et al., 2001; Cutler et al., 2001; Zwitserlood, 2003, 2004). Interestingly, word-initial syllables showed a more pronounced N100 in the EEG, compared to word-medial or final syllables (Sanders and Neville, 2003), and a study with newly learned pseudowords showed a similar N100 effect (Sanders et al., 2002).

Early evidence for the role of syllables in speech perception has been obtained in studies using monitoring paradigms (e.g., Mehler et al., 1981). In these studies, a (visually or auditorily presented) fragment precedes a spoken word, and participants have to press a button whenever the fragment is contained in the word (e.g., PA/PAL in the French word “palace”). In this paradigm, known as fragment or sequence monitoring (cf. Frauenfelder and Kearns, 1996), reactions are often faster when the fragment corresponds to the first syllable of the spoken word (e.g., PA – /palace/) than when not (e.g., PAL in /palace/). This has been taken as



evidence that listeners syllabify the incoming speech signal, and for syllable-based mental representations that mediate access to the lexicon (cf. Bradley et al., 1993). Using fragment-monitoring paradigms, syllable effects have been demonstrated in various languages, such as French, Spanish, Italian, Dutch, and Portuguese (Mehler et al., 1981; Morais et al., 1989; Bradley et al., 1993; Zwitserlood et al., 1993; Dumay and Content, 2012; see Floccia et al., 2012, for an excellent overview). The evidence is mixed for English (Cutler et al., 1986; Bradley et al., 1993).

Syllables do not seem to be good candidates for prelexical processing across all languages, and there is good reason to expect varying results in different languages. Languages differ greatly in the extent to which syllable boundaries are clear and unambiguous, as well as in the regularity of syllable structure (Bradley et al., 1993). English includes many ambisyllabic segments (i.e., segments that are part of two adjacent syllables). According to Bradley et al. (1993), a preponderance of ambisyllabicity may render syllabic segmentation inadequate. Consequently, other segmentation aids have been proposed (e.g., Cutler et al., 1986) for English listeners. Cutler et al. (1997) argue that listeners exploit the rhythmic structure characterizing their language in order to segment the speech signal. Cutler et al. (1997) assume that the viability of the syllable as an aid in speech segmentation corresponds to the basic prosodic structure of natural languages. In stress-timed languages such as English, listeners use stress as a segmentation cue, whereas in syllable-timed languages such as French and Spanish, listeners use syllabic boundary information. German and Dutch are stress-timed, but in Dutch, a language with high ambisyllabicity, syllabic effects have been observed (Zwitserlood et al., 1993). It is noteworthy that the interpretation that syllabic effects provide evidence for syllable-sized prelexical units of speech has been largely abandoned in the last decade or so. The more general view that syllable-boundary information—among various other cues—aids the parsing of speech input for lexical access has been proposed instead (Content et al., 2001; Cutler et al., 2001; Dumay et al., 2002; Zwitserlood, 2003, 2004).

Fragment or sequence monitoring is not the only paradigm with which effects of syllables in speech comprehension can be examined. An alternative is fragment priming, used to investigate lexical activation and access in spoken-word processing. Fragment priming can involve form relatedness (/kaep/ – captain) or semantic relatedness (/kaep/ – ship) between primes and targets (see Zwitserlood, 1989, 1996). The paradigm can be cross-modal, with spoken fragments and visually presented target words (similar to most fragment-monitoring studies) or unimodal, with spoken fragments and spoken targets. Form-related spoken fragments that match the target (e.g., /stri:/ – STREET) facilitate target processing relative to mismatching fragments (e.g., /stra:/ – STREET; cf. Marslen-Wilson, 1993). The paradigm has not been used often to investigate particular aspects of the fit between fragments and targets. Exceptions are the studies reported by Friedrich and colleagues. Friedrich et al. (2004a), for example, showed clear EEG correlates of segmental overlap between fragments and target words (e.g., /dra/ – DRAGON vs. /hun/ – DRAGON). They also investigated particular aspects of form overlap, for example shared place of articulation of

(differing) initial consonants of fragments and targets (Friedrich et al., 2008; Schild et al., 2012), or pitch (Friedrich et al., 2004b).

In the present study, we used fragment priming with lexical decision to study effects of syllabic match in German. Note that German has not been studied before with respect to a specific role for syllables in speech perception. There are studies in German on the role of syllables in visual word processing that demonstrate negative effects when the first syllable of a target word is of high frequency. This inhibition is evident in lexical decision (Conrad and Jacobs, 2004) and in early (200–300 ms) event-related components of EEG (Hutzler et al., 2004). The interpretation is that words are parsed into phonologically defined syllables during reading (cf. Conrad et al., 2007). However, interesting these results are, they provide no direct evidence for a similar role for syllables during speech processing. The fragment-priming studies with EEG by Friedrich and colleagues, all conducted in German, showed positive effects of segmental overlap. From the examples, the fragment primes seem to correspond to the first syllable of the target (e.g., /trep/ – TREPPE or /kan/ – KANTE), but it remains unclear whether there is a specific advantage of syllable-sized primes (Friedrich, 2005; Friedrich et al., 2009; Schild et al., 2012). Note also, that most studies employed a crossmodal paradigm (but see Friedrich et al., 2009; Schild et al., 2012), which seems less suitable to pick up early, or prelexical, effects of overlap between fragments and targets. We decided to use the unimodal priming variant (auditory fragments, auditory targets), because effects of syllabic match may be prelexical and/or modality-specific. We collected behavioral data—lexical decision latencies to spoken words and pseudowords primed by spoken fragments—and simultaneously recorded event-related potentials (ERPs).

Our participants performed lexical decisions on auditory stimuli (e.g., “lustig” – funny), preceded by related (e.g., /lus/) or unrelated (e.g., /tra/) auditory primes. Note that we used prime fragments spoken in isolation, not excised from longer stimuli. This was done to avoid particular information that is present in fragments that are cut out of words. One type of information comes from coarticulation of adjacent segments, another from subtle cues that signal syllabic boundaries. For example, Zwitserlood (2004) demonstrated for Dutch that fragments such as /mark/ cut out of /marker/ contain information about the syllabic boundary between /mar/ and /ker/. In fact, such cues drove the syllable-match effects obtained in that study. As our main aim was to study the role of syllables in speech processing at a level that abstracts away from particular cues provided in running speech, we opted for fragments spoken in isolation. Evidently, no solution is ideal, since fragments spoken in isolation tend to be longer than corresponding parts of longer words (Salverda et al., 2003). But note also that positive effects of overlap have been found before, with spoken prime-target stimuli that did not overlap completely (e.g., the French pseudoword “lurage” priming the target word “tirage”; Dumay et al., 2001).

Our predictions for the reaction time latencies were as follows. If related fragments activate corresponding words in the mental lexicon, lexical decision should be facilitated, compared to unrelated fragments. Crucially, if syllables play a role in German speech perception, related primes that precisely match the initial syllable, as in /lus/ – /lus.tig/ (funny), and /lust/ – /lust.los/

(listless; the dot marks the syllable boundary) should be superior to primes that match an equivalent number of initial phonemes but do not match the first syllable (e.g., /lus/ – /lust.los/, and /lust/ – /lus.tig/).

We also manipulated the relatedness between fragments and targets in the pseudoword trials. Note that this is hardly ever done, because pseudoword trials, necessary for the lexical decision task, are often considered uninteresting for other purposes (but see Friedrich et al., 2004a). Our pseudoword trials differed in critical aspects from the word trials. First, the fragments used with pseudoword targets did not correspond to existing German words or morphemes (e.g., wos, zas, limp, wost). The idea was to assess, with the pseudoword sets, effects of segmental overlap—even of syllable-sized segmental overlap—under conditions that minimize lexical contributions to the effects. For the same reason, the pseudoword targets were not very similar to existing words (e.g., wosteck, limpal, zastig) and many of the fragment primes did not even correspond to existing syllables (in particular the long primes). We explicitly avoided pseudowords consisting of two existing morphemes, such as “lustbar” or “mutung” that could exist but happen not to occur in the language (see Bölte et al., 2009, for EEG data on such stimuli). Thus, the pseudoword stimuli, in addition to their purpose for lexical decision, were used to assess the contribution of (syllable-sized) form overlap to the processing of spoken stimuli with as little lexical contribution as possible. Comparing word and pseudoword targets with respect to effects of (syllabic) match is informative with respect to the locus of these effects (lexical, prelexical; existing vs. possible syllables).

A comparison of behavioral and ERP data may shed light on the automaticity of potential effects and on their dependence on lexical processing, because EEG data are informative about the time course of effects. Based on the literature, we expected effects of relatedness in the EEG data. Both studies with auditory-auditory priming (Friedrich et al., 2009; Schild et al., 2012) revealed modulations of early (before 200 ms) components (N100, T-complex) by relatedness, as well as effects in the P350, a component sensitive to the goodness-of-fit between fragments and targets, reflecting lexical activation (Pykkänen and Marantz, 2003; Friedrich et al., 2009). Effects in the N400 range are taken to reflect lexical processing, that is, the fit between the input provided by the prime fragment and the lexical representation of the spoken target. As for syllabic match, there are no data to guide our predictions. Early modulations of the EEG should be evident when syllabic match plays a role during early phases of speech processing and lexical access. Modulations in the N400 domain would rather point to late, lexical effects, so a difference between word and pseudowords is expected here.

METHODS

PARTICIPANTS

Seventeen students of the Westfälische-Wilhelms Universität Münster, Münster, Germany (four males) with mean age 21 years (*SD* = 3.4, aged 19–31 years) took part in the experiment. All participants were native speakers of German and right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). None reported any (history of) hearing loss or

neurological problems. They received 12€ or course credit for their participation.

MATERIALS

All stimuli were spoken by a trained female native speaker of German and recorded using a high-quality microphone and a digital recorder (M-AUDIO microtrack 24/96) with a sampling rate of 44.1 Hz. For stimulus extraction and editing, we used the software packages *Praat* (Boersma and Weenink, 2010; version 5.0.23) and *CoolEdit* (CoolEdit 2000 v1.1).

The experiment contained 37 pairs of word targets, all of which were bisyllabic with clear syllable boundaries (e.g., lus.tig<sup>1</sup>, funny, and lust.los, dull). All word targets were morphologically complex, derived words. The targets of each pair differed in length (Long vs. Short; for durations and other relevant matching parameters see Appendix 1 in Supplementary Material), shared the initial morpheme (“lust,” delight), but differed in the length of their first syllables, due to re-syllabification of these morphemes (e.g., lust.los vs. lus.tig). As shown in **Table 1**, Short and Long targets were combined with two form-related (e.g., /lus/, /lust/) and two unrelated (e.g., /tra/, /trag/) spoken primes. Prime fragments were not excised from target words but recorded separately as monosyllabic stimuli. The two targets of a pair (e.g., lust.los – lus.tig) were combined with the same four primes, forming a set of eight trials with which stimulus properties and priming effects can be disentangled. In half of the trials, prime and target were related (e.g., /lus/ – /lust.los/), in half, they were unrelated (e.g., /trag/ – /lust.los/). Given that primes in the two related conditions differed in length, related, and unrelated primes were matched for length (/lus/ – /tra/; /lust/ – /trag/; related primes were some 20 ms shorter than unrelated ones), as well as with for pitch and intensity (see Appendix 1 in Supplementary Material). Unrelated primes did not occur as related primes elsewhere.

Crossed with the factor Relatedness was the factor Syllabic Match, which is a dummy variable for unrelated trials (see below). In half of all trials, indicated by a “+” sign in **Table 1**, the structure and segments of the prime matched the target-initial syllable (e.g., /lus/ – /lus.tig/, funny). In the other half, indicated by a “–” sign, despite shared phonemes, there was no syllable-structure

Table 1 | Example of materials.

Lexicality	Syllabic match	Unrelated	Related	Target length	Target
Word	+	/trag/	/lust/	Long	lust.los
	–	/tra/	/lus/		lust.los
	+	/tra/	/lus/	Short	lus.tig
	–	/trag/	/lust/		lus.tig
Pseudoword	+	/womp/	/bent/	Long	* bent.lok
	–	/wom/	/ben/		* bent.lok
	+	/wom/	/ben/	Short	* ben.to
	–	/womp/	/bent/		* ben.to

Note: A dot marks a syllable boundary. An asterisk marks a pseudoword.

<sup>1</sup>The dot marks a syllable boundary.

match (e.g., /lus/ – /lust.los/, listless). In related trials, the target always contained the prime. Only in half of the related trials did the prime exactly match the initial syllable of the target (e.g., /lus/ – /lus.tig/; /lust/ – /lust.los/). In the remaining half, the related prime was either shorter (e.g., /lus/ – /lust.los/) or longer (e.g., /lust/ – /lus.tig/) than the initial syllable. As is standard in control trials, the unrelated primes were in all aspects unrelated to their targets. With very few exceptions, this also concerned the syllabic skeleton of control primes and targets, so that the syllabic structure (CV, CVC, CCV, CCVC, and so on) of unrelated and related primes differed. Appendix 4 in Supplementary Material contains all word and pseudoword stimuli.

While Target Length and Syllabic Match are crossed for related pairs, there remains an imbalance with respect to the morphological status of the prime fragments. The longer prime fragments always corresponded to the stem morpheme of both target words. All stimuli had a transparent semantic relation between the stem morpheme and both targets. Given that many of the short targets were verbs, the long primes constituted the verbal stem of these stimuli (e.g., /greif/ in /greifen/, to grasp) and were potentially even more closely related to these verbs than to the derived longer words (e.g., /greifbar/ – graspable). Note also that in some cases, the short, CVC prime fragments corresponded to an existing morpheme (e.g., /lau/, tepid, from the target pair /lauf.band/ – /lau.fen/, treadmill, walk), but these morphemes were always semantically unrelated to the targets.

Another 37 sets of pseudowords (non-existing but phonotactically legal strings) were added for the lexical decision task. These sets occurred in similar conditions as the word sets. The pseudoword targets were created to be dissimilar to existing words while conforming to the phonotactic constraints of German. Almost without exception (e.g., /ris/ – Riss, crack), the fragment primes of pseudoword targets did not correspond to existing words or root morphemes of German. Note that syllabic overlap in fragment–pseudoword pairs is defined via the syllabic structure of the pseudoword targets and often involves non-existing, but possible syllables of German.

With 74 experimental sets (37 word sets, 37 pseudoword sets), and eight prime-target pairings per set, the number of experimental trials presented to each participant was 592. These trials were distributed over four blocks, such that there was no repetition of a prime or a target within a block. Using Latin square designs, conditions were evenly distributed over blocks, and block order was balanced between participants. To reduce the proportion of trials with prime-target overlap, we added 144 filler trials with 72 filler targets (36 words, 36 pseudowords) each combined with two different, unrelated prime fragments. None of these filler primes

or targets was used in an experimental trial. The experiment started with seven additional warm-up trials of similar structure.

# PROCEDURE

Participants were individually tested, comfortably seated in front of a computer screen (Samsung SyncMaster 2233RZ, 22", 120 Hz refresh rate, 1680 × 1050 pixel, 32 bit color depth) and a button box (Response Pad, Model RB – 830, producer Cedrus Corporation). They were instructed before and kept informed during the testing phase. Upon informed consent, the EEG-cap was positioned on the participant's head and two researchers simultaneously prepared the 64 electrodes. The experiment was controlled using the software *Presentation* (producer Neurobehavioral Systems, version 14.1.). To minimize artifacts, we asked participants to keep looking at a fixation cross at the center of the screen, and to blink and move as little as possible during trials. The auditory stimuli were presented via Sennheiser IE6 in-ear headphones and participants were allowed to adjust the volume to their individual preferences.

At the beginning of each trial (see **Figure 1**), a black fixation cross (Courier New, 48 pt) was presented at the center of the white screen where it remained until the end of the trial. Four hundred milliseconds after the appearance of the fixation cross, the auditory prime was presented, followed with a temporal jitter of 275–300 ms by the auditory target. Participants could provide their lexical decision from target onset onwards, pressing one of two buttons using their index fingers. For individual participants, left-right button assignment to word and pseudoword decisions remained the same throughout the experiment, between participants it was balanced. Participants were instructed to decide as quickly and accurately as possible. Decisions and the reaction times were recorded starting from target onset.

We followed standard EEG recording and analysis procedures (Picton et al., 2000). The EEG was recorded continuously from 64 Ag/AgCl electrodes using a WaveGuard cap (ANT Software B.V., The Netherlands) connected to a high input impedance amplifier (ANT ASA-lab amplifier, digital low-pass FIR-filter, cut-off frequency = 0.27 \*sampling rate). Two additional electrodes were placed on the outer left and right canthi and another two above and below the left eye to monitor eye movements. Impedances were kept below 10 kΩ. A high-impedance amplifier in combination with actively shielded electrode caps enables clear signals even with high electrode impedances (Ferree et al., 2001). The EEG was recorded with a sampling rate of 256 Hz using an average reference (Dien, 1998). Triggers were set to the beginning of the target stimulus.

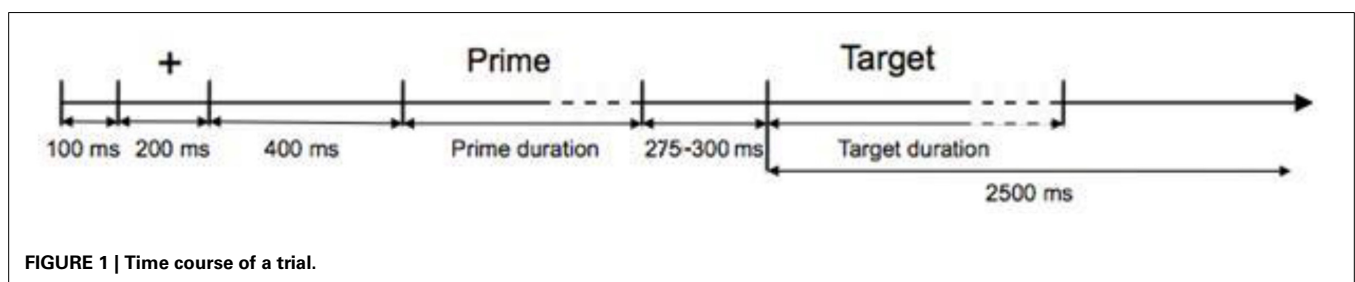


FIGURE 1 | Time course of a trial.

Every 1.5–2 min there was a short break of 10–15 s. Every 7.5–9 min there was a longer break of 1.5–2 min. During breaks, participants were allowed to move freely. The breaks were counted down by seconds on the screen, enabling participants to resume a comfortable position before the start of the consecutive trials. With 743 trials and the various breaks, the experiment lasted approximately 70 min. Including instructions, application and removal of the EEG-cap, a session took about 2 h.

## RESULTS

### BEHAVIORAL RESULTS

Trials with reaction times (RTs) above 2500 ms (0.2%) were excluded as were trials with incorrect lexical decision (1.6%). Error rates were below 5% for all participants. Three of the 74 sets were excluded from analyses due to high error rates for one of the targets. For the remaining 71 sets, we performed an outlier correction, excluding RTs outside two standard deviations from the mean per participant and condition (5.5%). The excluded trials as well as the remaining errors were evenly distributed over conditions and not analyzed further.

Overall, mean RTs were shorter for word (963 ms) than for pseudoword targets (1036 ms). This is a common finding in lexical decision and indicates that the pseudowords were not easy to reject as existing words (compared to phonotactically illegal stimuli such as “prlaspkusx”) even though they turned into pseudowords well-before word offset (see Table 2 for means and SD per condition). We ran Three-Way repeated measures ANOVAs with Relatedness (related vs. unrelated), Syllabic Match (present/absent in related conditions) and Target Length (short vs. long), separately for reaction times to word and pseudoword targets. The ANOVA using RT toward words revealed a significant effect of Target Length [ $F_{(1, 16)} = 10.200$ ,  $p = 0.006$ ,  $\eta_p^2 = 0.389$ ]. Not surprisingly, short targets (mean: 953 ms, SE: 35) yielded faster RTs than long targets (mean 972 ms, SE: 35). The interaction of Relatedness and Syllabic Match proved to be significant [ $F_{(1, 16)} = 27.098$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.629$ ]. This interaction showed that reactions to targets were facilitated when the related prime matched their first syllable (25 ms), but not when preceded by a prime that merely shared initial segments with the target (–13 ms). The three-way interaction [ $F_{(1, 16)} = 7.969$ ,

$p = 0.012$ ,  $\eta_p^2 = 0.332$ ] was also significant. All other effects were not significant ( $F = 1.481$ ,  $p = 0.214$ ,  $\eta_p^2 = 0.085$ ).

The three-way interaction was evaluated by means of *t*-tests, contrasting mean RTs to targets after related and unrelated primes. There was a clear 35 ms priming effect of short primes followed by their matching targets [e.g., /luf/ – /luf.tig/,  $t_{(16)} = 2.606$ ,  $p = 0.009$ ]. The smaller effect (13 ms) for long primes and their syllable-matching targets (e.g., /luft/ – /luft.los/) failed significance [ $t_{(16)} = 1.500$ ,  $p = 0.076$ ]. There was no facilitation of lexical decision latencies in cases of mere phonological overlap. When the fragment primes matched the segments of the target but not its first syllable (e.g., /luft/ – /luf.tig/) the short targets revealed a numerical 24 ms interference effect which was not significant [ $t_{(16)} = 1.634$ ,  $p = 0.061$ ] despite the fact that the long fragment prime corresponds to the stem morpheme. The condition with long targets (e.g., /luf/ – /luft.los/) showed no effect (–2 ms,  $t < 1$ ).

The ANOVA using RT toward pseudowords yielded a different pattern. Only the factors Relatedness [ $F_{(1, 16)} = 5.157$ ,  $p = 0.037$ ,  $\eta_p^2 = 0.244$ ] and Target Length [ $F_{(1, 16)} = 44.304$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.735$ ] proved significant. Short targets (mean: 1013, SE: 39) attracted faster RTs than long targets (mean: 1059, SE: 43). Pseudoword targets were responded to faster when preceded by unrelated primes (mean: 1029, SE: 42) than by related ones (mean 1043 ms, SE: 40). All other effects were not significant ( $F = 2.736$ ,  $p = 0.118$ ,  $\eta_p^2 = 0.146$ ).

To summarize the reaction-time data: No facilitation was evident for pseudowords preceded by related fragments; inhibition was observed instead. In contrast, there was facilitation for word targets preceded by fragments that matched their first syllable. This main effect of syllabic match was strongly evident for short fragments and short targets, and just failed significance for long fragments/long targets. This is surprising, given that the matching fragments (e.g., lust) of long targets (e.g., lust.los) correspond to the syllable as well as to the stem morpheme of the target. In Dutch, this double overlap resulted in larger effects than mere syllabic overlap evident with short prime fragments, but note that these data come from fragment monitoring, not from priming (Zwitserslood, 2003). In contrast to syllabic match between fragments and targets, the cases of syllabic mismatch—but still providing phonemic overlap—showed no priming.

The difference between word and pseudoword targets can be interpreted in two ways. First, it is possible that the existence of the primes as syllables of the language, for example as members of a mental syllabary (Levelt and Wheeldon, 1994), is a prerequisite for priming by syllable-sized fragments. If this plays a role, we would have expected a difference between the long and short pseudoword primes, since about half of the short primes were (rather infrequent) syllables of German, but the long ones were not.

An intriguing finding is the lack of priming with long fragments and short word targets, when related primes mismatch the first syllable but still constitute targets’ stem morpheme. In fact, this condition with syllabic mismatch showed numerical interference instead of facilitation. In contrast, the pseudoword conditions—in the first-block ANOVA—showed interference in cases of syllabic match, that is, in conditions that produce

**Table 2 | Mean reaction times in ms and SD (in parentheses) as a function of conditions.**

Lexicality	Relatedness	Syllabic match	Target length	
			Short	Long
Word	Related	+	930 (132)	967 (153)
		–	971 (147)	972 (142)
	Unrelated	(+)	965 (144)	980 (146)
		(–)	947 (164)	970 (145)
Pseudoword	Related	+	1023 (155)	1069 (183)
		–	1013 (153)	1066 (175)
	Unrelated	(+)	1016 (167)	1046 (190)
		(–)	999 (176)	1054 (173)



facilitation for word targets. A possible interpretation for this reversal of effects focuses on the nature of the targets, word or pseudowords. The only positive syllabic effects occurred with word targets, interference was observed when the target was a pseudoword, in particular for related primes that syllabically matched their target. This suggests a sensitivity to the syllabic structure of the targets that results in speeded word decisions and slowed pseudoword decisions. A likely locus for such a pattern is a postlexical, strategic one.

Auditory lexical decision thus seems to tap into late effects of syllabic match and may not be ideally suited to investigate the role of syllables in pre-lexical and early lexical speech processing. For this, EEG data might be better suited.

## ERP RESULTS

EEG-data were analyzed using a combination of ASA (ANT, The Netherlands), EEGLAB (version 12.0.2.06b, Delorme and Makeig, 2004; MATLAB 2012b) and ERPLAB (version 4.0.2.3, Lopez-Calderon and Luck, 2014). EEG-data were filtered using a half-power Butterworth bandpass filter (0.1–20 Hz, 24 dB/oct) based on the FFT-method. Ocular artifacts were corrected using a PCA-approach (Ille et al., 2002). Remaining artifacts were detected using a  $\pm 75 \mu\text{V}$  threshold. There were on average 11% errors in the word condition and 10% errors in the pseudoword condition (see Appendix 2 in Supplementary Material for a complete compilation). Artifact-free trials with correct responses were averaged using epochs of 700 ms length, time locked to target onset, with a 200 ms pre-stimulus baseline. We formed six regions of interest (anterior central: Fp1, Fp2, Fp3, AF3, AF4, F1, Fz, F2; anterior left: F7, F5, FT7, FC5, FC3, T7, C5, C3; anterior right: F8, F6, FT8, FC6, FC4, T8, C6, C4; posterior central: CP1, CP2, CP3, Pz, P4, POz, Cz; posterior left: TP7, CP5, P7, P5, PO7, PO5, PO3, O1; posterior right: TP8, CP6, P8, P6, PO8, PO6, PO4, O2). These six ROIs constituted the variables LR-axis (left, central, right) and A-P (anterior, posterior).

The EEG-data quality for two participants was too low to be included in further analyses. Mean voltage was calculated in a number of time windows that were shown to be of interest in unimodal fragment priming (cf. Friedrich et al., 2009; Schild et al., 2012). The first two windows (80–200 ms, including the N100) and (200–300 ms) are taken to reflect early modality-specific processing of speech input (cf. Friedrich et al., 2009). Next, data were analyzed in time windows that were shown to be relevant for auditory-auditory priming with matching or mismatching word fragments: the 300–400 ms window, including potential P350 effects, and the 280–500 ms window, including the N400, where lexical effects are expected next to effects of the overlap between fragments and targets.

EEG data for correct reactions in lexical decision were included, and all ANOVAs (except the last one) had the following within-factors: Syllabic Match (present vs. absent in related conditions), Relatedness (related vs. unrelated), Target Length (short, long), A-P (anterior, posterior), and LR-axis (left, central, right). The ANOVA on the N400 time window included a different electrode selection. Word and pseudoword targets were analyzed separately. Effects including more than two levels are reported only if they remained significant after Greenhouse-Geisser correction; effects of electrodes (A-P,

LR-axis) are reported only when interacting with manipulated factors.

## TIME WINDOW 100–200 MS

The ANOVAs showed no significant effects of any of the variables, nor interactions between them, in this time window. This held for word and pseudoword targets alike.

## TIME WINDOW 200–300 MS

The repeated-measure ANOVA with Syllabic Match, Relatedness, Target Length, A-P, and LR-axis on the mean amplitudes for words in the time window from 200 to 300 ms showed a significant interaction between Relatedness and A-P [ $F_{(1, 14)} = 15.76$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.529$ ] as well as a three-way interaction between Relatedness, A-P and LR-axis [ $F_{(2, 28)} = 5.45$ ,  $p = 0.009$ ,  $\eta_p^2 = 0.280$ ,  $GG = 0.996$ ; see Figure 2]. Related conditions showed more negative mean amplitudes at anterior sites than unrelated conditions, and this pattern was reversed at posterior sites. The three-way interaction revealed that the difference in  $\mu\text{V}$  between related and unrelated conditions was most pronounced at left-anterior electrodes (0.76). Comparing related and unrelated conditions, all differences were significant (Fischer's LSD = 0.206) except for the anterior-central (0.10) electrodes (see Figure 2). The localisation and the polarity of this effect fits best with a P350.

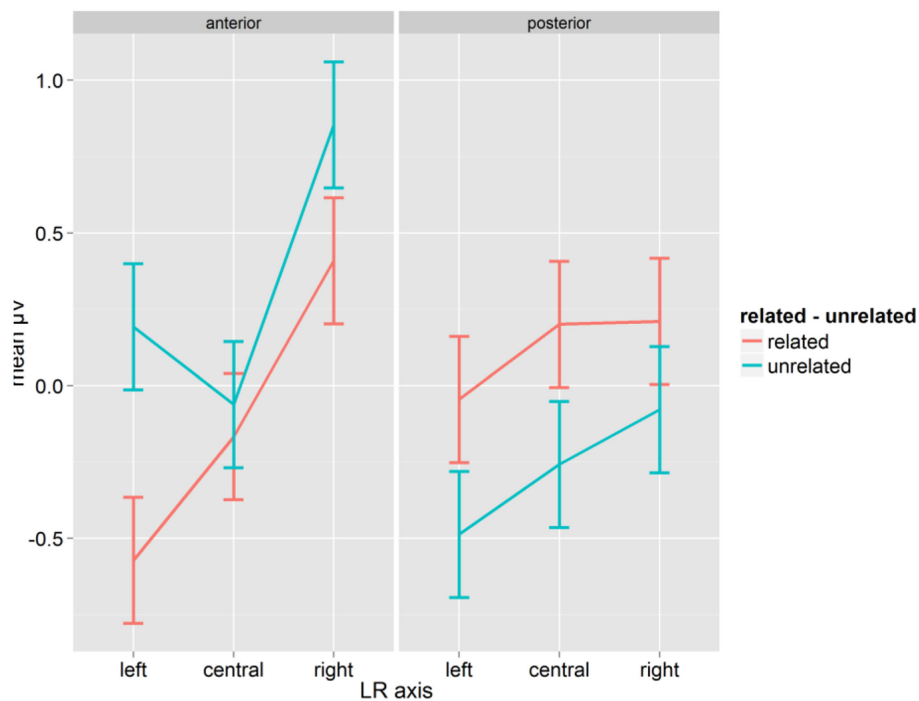
The analysis also showed a main effect of Syllabic Match [ $F_{(1, 14)} = 8.20$ ,  $p = 0.013$ ,  $\eta_p^2 = 0.369$ ]. Collapsed over related and unrelated primes, “matching” conditions yielded more positive mean amplitudes (mean:  $0.048 \mu\text{V}$ , SD: 0.929) than “mismatching” conditions (mean:  $-0.015 \mu\text{V}$ , SD: 0.973). Given that Syllabic Match only applies to related primes, and despite the fact that the interaction between Syllabic Match and Relatedness was not significant [ $F_{(1, 14)} = 1.90$ ,  $p = 0.189$ ], we computed separate ANOVAs on related and unrelated prime-target pairs. The ANOVA on related prime-target pairs showed an effect of Syllabic Match [ $F_{(1, 14)} = 6.88$ ,  $p = 0.020$ ,  $\eta_p^2 = 0.330$ ]. Matching prime target pairs (mean:  $0.053$ , SD: 0.150) were more positive than mismatching ones (mean:  $-0.041$ , SD: 0.106). Although in the same direction, ( $0.042 \mu\text{V}$  for “matching” pairs,  $0.011 \mu\text{V}$  for “mismatching” pairs), the effect failed significance in the ANOVA on unrelated prime-target pairs [ $F_{(1, 14)} = 1.23$ ,  $p = 0.285$ ].

The analysis of pseudoword reactions showed no effects or interactions of Relatedness and Syllabic Match, neither overall nor in separate ANOVAs on related and unrelated prime-target pairs (all  $p > 0.12$ ). However, the interaction of Relatedness, A-P and LR-axis was significant [ $F_{(2, 28)} = 15.11$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.519$ ,  $GG = 0.889$ , see Figure 3], with a very similar pattern to the P350 found for words. Fischer's LSD for the contrast of related and unrelated conditions was 0.153, and differences were significant except at anterior and posterior central sites (see Figure 3).

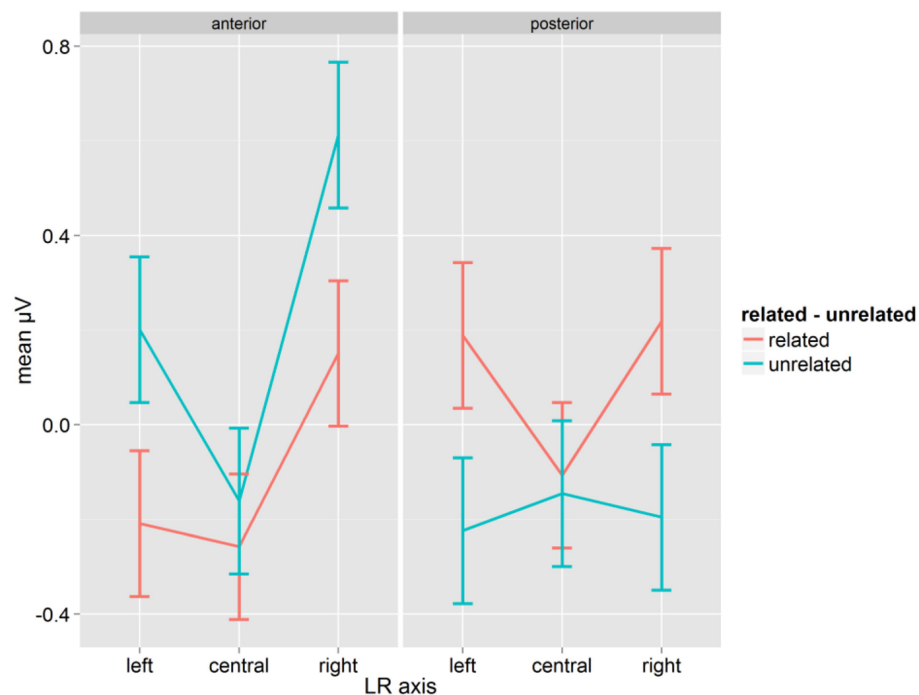
## TIME WINDOW 300–400 MS

The same ANOVA as before was run for correct word responses using the mean amplitude in a time window of 300–400 ms. There was a significant main effect Relatedness [ $F_{(1, 14)} = 5.22$ ,  $p = 0.038$ ,  $\eta_p^2 = 0.271$ ]. Overall, the amplitude of the related condition (mean:  $0.055$ , SD: 1.14) was less positive than the amplitude of the unrelated condition (mean:  $0.094$ , SD: 1.10).





**FIGURE 2 |** Mean  $\mu V$  for words as a function of conditions for the time-window 200–300 ms; error bars represent Fisher's least significant differences.



**FIGURE 3 |** Mean  $\mu V$  for pseudowords as a function of conditions for the time-window 200–300 ms; error bars represent Fisher's least significant differences.

Relatedness interacted with A-P and with LR-axis, and the three-way interaction between Relatedness, A-P and LR-axis was also significant [ $F_{(2, 28)} = 4.96$ ,  $p = 0.014$ ,  $\eta_p^2 = 0.261$ ,  $GG = 0.912$ ; see **Figure 4**]. The mean amplitudes in the related condition were more negative than in the unrelated condition at anterior sites, and this pattern was reversed at posterior sites. The effect (related–unrelated) was most pronounced at left-anterior and central-posterior sites (0.94 for both), and much smaller (or insignificant, see **Figure 4**; Fischer's LSD = 0.225) at other sites. As with the earlier window, this pattern corresponds best with a P350. There was also a clear reversed effect (unrelated more negative than related) at posterior central electrodes, which seems indicative of an N400.

Collapsed over related and unrelated conditions, there was a main effect of Syllabic Match [ $F_{(1, 14)} = 4.95$ ,  $p = 0.043$ ,  $\eta_p^2 = 0.261$ ; match mean: 0.095, SD: 1.14; mismatch mean: 0.053, SD: 1.10]. As for the 200–300 time window, we analyzed related and unrelated prime-target pairs separately, because Syllabic Match is a dummy variable for unrelated trials. The effect of Syllabic Match, with matching targets showing a more positive mean amplitude than mismatching targets (match mean: 0.09, mismatch mean: 0.01) just failed significance in the ANOVA on related prime-target pairs [ $F_{(1, 14)} = 4.42$ ,  $p = 0.054$ ,  $\eta_p^2 = 0.240$ ]. Syllabic Match was not significant for unrelated prime-target pairs ( $F < 1$ ).

The ANOVA on pseudowords yielded a significant interaction between Relatedness and A-P [ $F_{(2, 28)} = 17.09$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.549$ ], but the three-way interaction with LR-axis was not significant. As with words, mean amplitudes were more negative in related (−0.40) than in unrelated (0.01) conditions at anterior

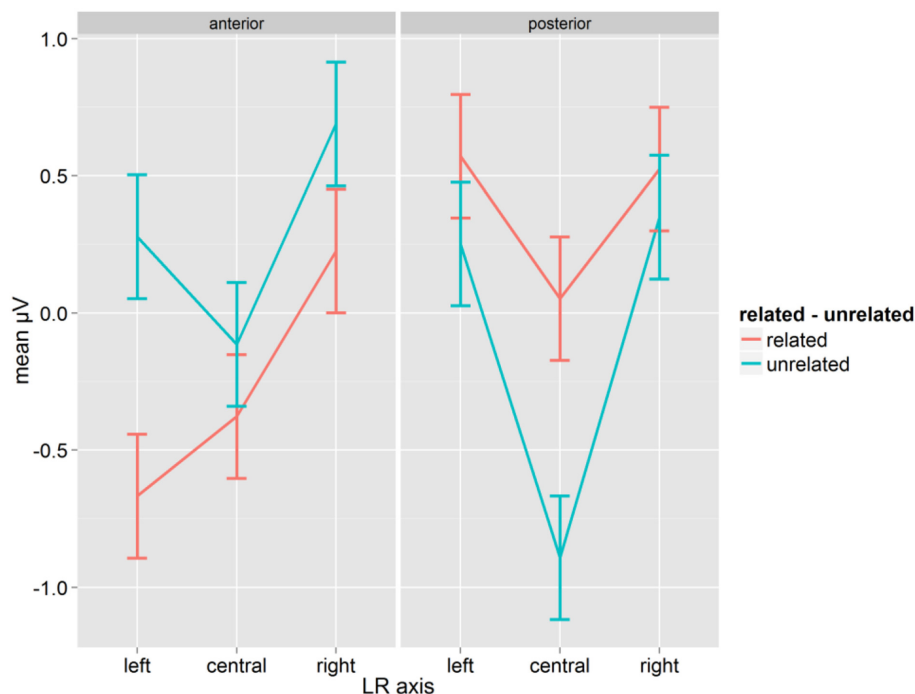
regions, while the reverse (related: 0.52; unrelated 0.16) was true at posterior regions (both effects are significant; Fischer's LSD = 0.277). As for the earlier time window, the polarity fits well with an anterior P350 effect. There were no effects of Syllabic Match for pseudoword targets in this time window, not in the overall analysis, nor in the analyses on related and unrelated trials separately (all  $F < 1$ ).

#### TIME WINDOW 280–500 MS

Visual inspection showed a N400 at central-posterior electrodes, when comparing word targets preceded by related and unrelated prime fragments, with and without Syllabic Match (see Appendix 3: Figures 1, 2 in Supplementary Material). Based on the literature, we calculated the mean amplitude of a region of interest consisting of the electrodes C1, Cz, C2, CP1, CPz, CP2, P1, Pz, and P2 in a time window of 280–500 ms<sup>2</sup>. This served as the dependent measure in a Three-Way repeated measurement ANOVA with the factors Relatedness (related vs. unrelated), Lexicality (word vs. pseudoword), and Syllabic Match (present vs. absent in related conditions). No Greenhouse-Geisser-correction was needed because all factors had only two levels.

The analysis revealed a significant main effect of Relatedness [ $F_{(1, 14)} = 37.24$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.727$ ]. The mean amplitude in unrelated conditions was more negative (mean: −1.034, SD:

<sup>2</sup>Using a time-window of 350 ms to 450 ms yielded the same significant main effect of Relatedness [ $F_{(1, 14)} = 40.949$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.745$ ] and the significant interaction of Relatedness and Lexicality [ $F_{(1, 14)} = 12.599$ ,  $p = 0.003$ ,  $\eta_p^2 = 0.474$ ]. The interaction of Syllabic Match and Lexicality [ $F_{(1, 14)} = 2.745$ ,  $p = 0.120$ ,  $\eta_p^2 = 0.164$ ], however, was not significant.



**FIGURE 4 |** Mean  $\mu V$  for words as a function of conditions for the time-window 300–400 ms; error bars represent Fisher's least significant differences.

0.686) than in related conditions (mean:  $-0.464$ , SD:  $0.745$ ). Importantly, the interaction of Relatedness and Lexicality was also significant [ $F_{(1, 14)} = 12.51$ ,  $p = 0.003$ ,  $\eta_p^2 = 0.472$ ]. With Fischer's LSD of  $0.23$ , related words (mean:  $-0.293$ , SD:  $0.746$ ) and related pseudowords (mean:  $-0.635$ , SD:  $0.803$ ) were reliably less negative than their unrelated counterparts (words mean:  $-1.133$ , SD:  $0.900$ ; pseudowords mean:  $-0.936$ , SD:  $0.538$ ). However, this difference—the N400 effect—was almost four times as large for words as for pseudowords. Furthermore, the interaction of Syllabic Match and Lexicality [ $F_{(1, 14)} = 5.37$ ,  $p = 0.036$ ,  $\eta_p^2 = 0.277$ ] was significant. Collapsed over Relatedness, amplitudes to words were more negative with “matching” than with “mismatching” syllables, and this was reversed for pseudowords. Given that these data include the control trials to which syllabic match does not apply, separate analyses were calculated for related and unrelated prime-target pairs, for words and pseudowords separately. There were no effects of syllabic match, neither for related nor for unrelated prime-target pairs.

To summarize the ERP data: except for the earliest time windows (100–200 ms), there is clear evidence for effects of the relatedness between fragment primes and targets. We also observed main effects of syllabic match. Given that control primes neither share segments with, nor match the (abstract) syllabic structure of, the targets, main effects of syllabic match most probably reflect a correspondence in length of fragments and targets (long/long and short/short < long/short and short/long)—independent of phonological relatedness. However, such effects do not elucidate the impact of segmental and/or syllabic overlap between related primes and targets. Of interest for this question are main effects of relatedness between primes and targets, and interactions between relatedness and the other factors—most importantly syllabic match. Despite the fact that we found no significant interactions in our EEG data, we performed separate analyses on related and unrelated prime-target pairs. In the 200–300 ms window, there was an effect in related conditions. Targets whose first syllable corresponded to the prime fragment (e.g., /lus/ – /lus.tig/) elicited more positive values than targets preceded by fragments that mismatched their first syllable (e.g., /lus/ – /lust.los/). There was a similar trend in the 300–400 ms window. Unrelated pairs and pseudoword conditions showed no such effects in any time window.

With respect to the overall impact of related primes, a very similar pattern was observed in the two time windows lasting from 200–300 to 300–400. Related conditions showed more negative amplitudes than unrelated ones at anterior sites, but more positive amplitudes than unrelated ones at posterior sites. The strongest effects were observed over left-anterior electrodes. Clearly, this is not an N400 type effect. Given its left-anterior dominance, time window (200–400 ms), and polarity, the observed effect fits best with a modulation of the P350. Interestingly, these P350 effects were not qualified by an interaction with syllabic match. Thus, these differences between related and unrelated conditions held for all cases of overlap—syllabic or not. Note also that very similar patterns were observed for word and pseudoword targets in these time windows. The 300–400 ms window also revealed a pronounced central negativity for word trials, with more negative amplitudes for unrelated than

for related conditions. This is similar to the central negativity observed by others (Friedrich et al., 2009; Schild et al., 2012). Given that the 300–400 time window largely overlaps with the time window from 280 to 500, we believe that this central negativity in fact corresponds to an N400.

The analysis on the 280–500 ms time window, on a selection of electrodes often used for N400-analyses, showed the expected difference between related and unrelated targets in the N400 time window. The N400 was more negative in unrelated than in related fragment conditions. This was qualified by an interaction, showing that this N400 effect was much larger for words than for pseudowords. No effects of syllabic match were observed in the N400 time window.

None of the EEG data revealed effects of syllabic match in interaction with relatedness. If related fragment primes that match the first syllable of the target had a special status, this should have revealed itself in such an interaction, because the control primes did not match the (abstract) syllabic structure. The only reliable evidence for an impact of syllabic match in related fragment-target pairs was observed in the 200–300 window; the 300–400 ms window showed a trend. Interesting as they are, these effects should be treated with caution since they did not reveal themselves in an interaction between syllabic match and relatedness, but in separate analyses of related and unrelated fragment conditions. Thus, whereas related prime fragments consistently have a different impact on ERP components than unrelated fragments from 200 ms onwards, this impact is not qualified by syllabic match, aside from the indication for related prime-target pairs in one time window.

## GENERAL DISCUSSION

This study used fragment priming, with related (e.g., /mu/) and unrelated (e.g., /tes/) spoken fragments to spoken targets words (e.g., /mu.tig/) or pseudowords. Relatedness was further varied along the dimension of syllabic match between fragments and targets. Whereas /mu/ specifies the initial segments of both /mutig/ and /mutlos/, it corresponds to the first syllable of /mutig/, not of /mutlos/. Syllabic match was implemented in related word and pseudoword conditions. Whereas obviously, all word targets were combined with existing syllables, the pseudoword targets that were phonotactically legal but not very word-like, were paired with fragments that, according to the rules of syllabification of German, structurally corresponded to their first syllables, but that in most cases were not part of the syllable inventory of the language.

First, we expected effects of relatedness, with an advantage (in RT) and differences (in EEG amplitude) between related and unrelated conditions. Second, if syllables play a role in German speech perception, related primes that precisely match the initial syllable should be superior to primes that match an equivalent number of initial phonemes but do not correspond to the first syllable. Third, differences in effects for words and pseudowords may inform us about the origin of effects of overlap. If words and pseudowords show similar effects, these may well originate from prelexical or early lexical levels of processing. If effects diverge, this indicates lexical involvement. Finally, a comparison of behavioral and ERP data, and of early and late EEG effects, may be

informative with respect to the automaticity of potential effects, and on their dependence on advanced lexical processing. What do the data tell us, and how do they compare to results from other studies, in particular from the—admittedly small number of—studies that use the same paradigm and measures? We discuss the behavioral and EEG data separately.

### REACTION TIMES IN AUDITORY LEXICAL DECISION

To start with the behavioral data: There are effects of overlap for both words and pseudowords. For pseudowords, segmental overlap slows down lexical decision, for words, overlap speeds up reactions. Moreover, an effect of syllabic match of related fragments is present for existing words, most strongly in one particular condition. For pseudowords, overlap between fragments and targets slows down correct decisions. This may come as a surprise, since many fragments of pseudowords were not very wordlike by themselves. Note that their onsets (the first two or three segments), are compatible with existing words in the language and would thus activate lexical cohorts (Zwitserslood, 1989). In cases of segmental overlap between fragments and targets, this lexical activation may have interfered with a correct pseudoword decision on the targets. Moreover, the interference effect seems dependent on syllabic match between related primes and targets (in the analysis of the first-block data). This dependence on syllabic match of both the facilitation, with words, and the interference, with pseudowords, indicates a sensitivity to the syllabic structure of the targets that results in speeded word decisions and slowed pseudoword decisions. A likely locus for such a pattern is a postlexical, strategic one. Auditory lexical decision thus seems to tap into late effects of the syllabic match between prime fragments and their related targets, and may not be ideally suited to investigate the role of syllables during early phases of speech processing. The same has been argued for the monitoring task when it uses catch trials to prevent very fast decisions (see Zwitserslood, 2003; Floccia et al., 2012). In all, positive effects of syllabic match were not overly strong in the lexical decision task. Moreover, it is surprising that fragments that constituted the onsets of target words, but did not correspond to the first syllable, induced no facilitation at all (cf. Zwitserslood, 1989, 1996).

How do these behavioral data compare to those obtained by Friedrich and colleagues, in similar auditory-auditory fragment priming studies with lexical decision (Friedrich et al., 2009; Schild et al., 2012)? First, no data on pseudoword trials were reported in any of these studies. For word targets, both studies obtained significant facilitation, comparing latencies to targets after related and unrelated fragment primes. Taking a closer look at their materials, these fragments always corresponded to the first syllable of the spoken words. Thus, our behavioral data replicate effects reported for words in these studies, using the same paradigm to study the same language, German. Since both studies also registered EEG, it will be interesting to compare the ERP effects reported next.

### ERP DATA

The ERP data show no effects in the earliest time window (100–200 ms) that includes the N100. This is different from both studies that used auditory-auditory priming (Friedrich et al., 2009; Schild

et al., 2012), who reported N100 or T-complex/N100 effects as a function of relatedness between primes and targets in this window. We find no such effects, neither for words nor for pseudowords. Given that these early effects are generally not the most robust, this might be due to the smaller number of participants remaining in the EEG analyses in our study (15 vs. 22 in the other studies).

In the two consecutive windows (200–300 and 300–400 ms), we observe clear modulations of the P350. The polarity of effects as well as the affected electrodes fits with what is observed by others: in (left) anterior regions, related trials show a more negative P350 than unrelated trials (cf. Friedrich et al., 2004a, 2008; Pykkänen and Marantz, 2003, for the equivalent component from MEG). The P350 is taken to be sensitive to the degree of prime-target overlap. This fits well with the effects observed here, for both words and pseudowords. The P350 is also interpreted to reflect the activation of word-form representations. A related fragment facilitates lexical access relative to an unrelated fragment (see Friedrich et al., 2013). Note again that we obtain quite similar results for word and pseudowords targets. Given that pseudowords have no lexical representation, how can the P350 reflect facilitated lexical access? It should be noted that the first two, and often the first three, segments of the pseudoword primes and targets still correspond to existing words. Given the timing of the P350 and the moment in time at which spoken targets become pseudowords, it is quite feasible that their pseudoword status is not yet available to influence the phonological matching and word-form activation effects that are present and reliable throughout (see Friedrich et al., 2004a, for similar P350 effects with visual pseudoword targets).

In the N400 window, already present as a central negativity in the 300–400 ms window, the polarity of the relatedness effect reverses, with unrelated conditions being more negative than related conditions (see Appendix 3: Figures 1, 2 in Supplementary Material). This is in accordance with data for segmental overlap from many other studies (e.g., Praamstra et al., 1994; Dumay et al., 2001; Diaz and Swaab, 2007; Desroches et al., 2009; Scharinger and Felder, 2011). The N400 revealed an interaction between lexical status and relatedness. The N400, in terms of the difference between related and unrelated prime conditions, was much larger for words than for pseudowords—although the N400 for pseudowords was also reliable. In the N400 time window, lexical influences thus start to kick in, modulating the impact of form overlap between fragment primes and targets. Given that the time window extends to 500 ms, the information as to whether target stimuli are words or pseudowords should have become available for most stimuli. The pattern found for the N400 suggests that lexical selection is well on its way.

Given that we set out to investigate effects of syllabic match between fragment primes and targets, it is revealing that none of the EEG data revealed effects of syllabic match in interaction with relatedness. Only when—despite the lack of such interactions—related and unrelated conditions were analyzed separately did we observe an effect of syllabic match between prime fragments and word targets. As this is not statistically backed up by appropriate interactions, we feel somewhat reluctant to interpret this observation. Evidently, more research is needed to elucidate these

effects. In sum, whereas related prime fragments have a different impact on the ERP components than unrelated fragments in all but the earliest time windows, effects of syllabic match are ephemeral—and definitely absent in the N400 window. It is also noteworthy that the time windows prior to the N400 showed effects to be very similar for words and pseudowords, most probably because their lexical status is not yet clear. What the ERP data reveal, are the processes involved in phonetic/phonological matching, lexical access and selection during spoken-word recognition. The consistent advantage of related fragments for target processing, including the mapping of incoming speech, lexical access and selection, fits many models of spoken word recognition (McClelland and Elman, 1986; Marslen-Wilson, 1987; see also Zwitserlood, 1989). The obvious conclusion is that syllabic match is not crucial for these processes. Our data provide no support for syllable-sized prelexical representations that mediate between the speech input and the mental lexicon. Note that syllabic cues may still play an important role in speech segmentation; our fragment-priming paradigm does not really address this question (cf. Zwitserlood, 2003). Why, then, do lexical decision latencies show at least some effects of syllabic match? The most tempting interpretation is that syllabic effects obtained in behavioral data, such as the lexical decision latencies reported here, reflect either late lexical processing or even post-lexical strategic processing, but not speech perception and lexical access. This is supported by the fact that we found no evidence for syllabic match in reaction times to pseudowords. Note that the behavioral data are somewhat puzzling to start with, with no evidence of morphological priming between the longer fragments and both related targets. This pattern indicates a dissociation of more automatic processes—evident in the ERPs—and data from tasks that require conscious target processing—such as lexical decision. Such dissociations are an all-to-familiar phenomenon in research on early processes in speech perception (cf. Bien and Zwitserlood, 2013). With respect to the quest for the syllable as a prelexical unit of speech processing, the following quote elegantly sums up the problem: “In sum, although a lot of evidence indicates that the syllabic structure influences spoken word recognition, there is very little support for the idea that syllabic coding units are extracted from the signal and intervene in the perceptual processes” (Dumay and Content, 2012, p. 682).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.01544/abstract>

## REFERENCES

- Bien, H., Lagemann, L., Döbel, C., and Zwitserlood, P. (2009). Implicit and explicit categorization of speech sounds—dissociating behavioural and neurophysiological data. *Eur. J. Neurosci.* 30, 339–346. doi: 10.1111/j.1460-9568.2009.06826.x
- Bien, H., and Zwitserlood, P. (2013). Processing nasals with and without consecutive context phonemes: evidence from explicit categorization and the N100. *Front. Psychol.* 4:21. doi: 10.3389/fpsyg.2013.00021
- Boersma, P., and Weenink, D. (2010). *Praat: Doing Phonetics by Computer [Computer program]. Version 5.1.41.* Available online at: <http://www.praat.org/>
- Bölte, J., Jansma, B. M., Zilverstand, A., and Zwitserlood, P. (2009). Derivational morphology approached with event-related potentials. *Ment. Lex.* 4, 336–353. doi: 10.1075/ml.4.3.02bol
- Bradley, D. C., Sánchez-Casas, R. M., and García-Albea, J. E. (1993). The status of the syllable in the perception of Spanish and English. *Lang. Cogn. Process.* 8, 197–233. doi: 10.1080/01690969308406954
- Christiansen, M. H., and Chater, N. (2001). Connectionist psycholinguistics: capturing the empirical data. *Trends Cogn. Sci.* 5, 82–88. doi: 10.1016/S1364-6613(00)01600-4
- Christophe, A., Peperkamp, S., Pallier, C., Block, E., and Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *J. Mem. Lang.* 51, 523–547. doi: 10.1016/j.jml.2004.07.001
- Conrad, M., Grainger, J., and Jacobs, A. M. (2007). Phonology as the source of syllable frequency effects in visual word recognition: evidence from French. *Mem. Cognit.* 35, 974–983. doi: 10.3758/BF03193470
- Conrad, M., and Jacobs, A. M. (2004). Replicating syllable frequency effects in Spanish in German: one more challenge to computational models of visual word recognition. *Lang. Cogn. Process.* 19, 369–390. doi: 10.1080/01690960344000224
- Content, A., Kearns, R. K., and Frauenfelder, U. H. (2001). Boundaries versus onsets in syllabic segmentation. *J. Mem. Lang.* 45, 177–199. doi: 10.1006/jmla.2000.2775
- Cutler, A., Dahan, D., and van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Lang. Speech* 40, 141–201.
- Cutler, A., McQueen, J. M., Norris, D., and Somejuan, A. (2001). “The roll of the silly ball,” in *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*, ed E. Dupoux (Cambridge, MA: MIT Press), 181–194.
- Cutler, A., Mehler, J., Norris, D., and Segui, J. (1986). The syllable’s differing role in the segmentation of French and English. *J. Mem. Lang.* 25, 385–400. doi: 10.1016/0749-596X(86)90033-1
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Desroches, A. S., Newman, R., and Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: electrophysiological evidence for the influences of phonological similarity. *J. Cogn. Neurosci.* 21, 1893–1906. doi: 10.1162/jocn.2008.21142
- Diaz, M. T., and Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Res.* 1146, 85–100. doi: 10.1016/j.brainres.2006.07.034
- Dien, J. (1998). Issues in the application of the average reference: review, critiques, and recommendations. *Behav. Res. Methods Instrum. Comput.* 30, 34–43. doi: 10.3758/BF03209414
- Dumay, N., Benraiss, A., Barriol, B., Colin, C., Radeau, M., and Besson, M. (2001). Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words. *J. Cogn. Neurosci.* 13, 121–143. doi: 10.1162/089892901564117
- Dumay, N., and Content, A. (2012). Searching for syllabic coding units in speech perception. *J. Mem. Lang.* 66, 680–694. doi: 10.1016/j.jml.2012.03.001
- Dumay, N., Frauenfelder, U. H., and Content, A. (2002). The role of the syllable in lexical segmentation in French: word-spotting data. *Brain Lang.* 81, 144–161. doi: 10.1006/brln.2001.2513
- Ferree, T. C., Luu, P., Russell, G. S., and Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clin. Neurophysiol.* 112, 536–544. doi: 10.1016/S1388-2457(00)00533-2
- Floccia, C., Goslin, J., Morais, J. J., and Kolinsky, R. (2012). Syllable effects in a fragment-detection task in Italian listeners. *Front. Psychol.* 3:140. doi: 10.3389/fpsyg.2012.00140
- Frauenfelder, U. H., and Kearns, R. K. (1996). Sequence monitoring. *Lang. Cogn. Process.* 11, 665–673. doi: 10.1080/016909696387079
- Friedrich, C. K. (2005). Neurophysiological correlates of mismatch in lexical access. *BMC Neurosci.* 6:64. doi: 10.1186/1471-2202-6-64
- Friedrich, C. K., Felder, V., Lahiri, A., and Eulitz, C. (2013). Activation of words with phonological overlap. *Front. Psychol.* 4:556. doi: 10.3389/fpsyg.2013.00556
- Friedrich, C. K., Kotz, S. A., Friederici, A. D., and Alter, K. (2004b). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Cogn. Brain Res.* 20, 300–308. doi: 10.1016/j.cogbrainres.2004.03.007
- Friedrich, C. K., Kotz, S. A., Friederici, A. D., and Gunter, T. C. (2004a). ERPs reflect lexical identification in word fragment priming. *J. Cogn. Neurosci.* 16, 541–552. doi: 10.1162/089892904323057281



- Friedrich, C. K., Lahiri, A., and Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: asymmetries with word initial variations. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1545–1559. doi: 10.1037/a0012481
- Friedrich, C. K., Schild, U., and Röder, B. (2009). Electrophysiological indices of word fragment priming allow characterizing neural stages of speech recognition. *Biol. Psychol.* 80, 105–113. doi: 10.1016/j.biopsycho.2008.04.012
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Hutzel, F., Bergmann, J., Conrad, M., Kronbichler, M., Stenneken, P., and Jacobs, A. M. (2004). Inhibitory effects of first syllable-frequency in lexical decision: an event-related potential study. *Neurosci. Lett.* 372, 179–184. doi: 10.1016/j.neulet.2004.07.050
- Ille, N., Berg, P., and Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *J. Clin. Neurophysiol.* 19, 113–124. doi: 10.1097/00004691-200203000-00002
- Langus, A., Marchetto, E., Bion, R., and Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *J. Mem. Lang.* 66, 285–306. doi: 10.1016/j.jml.2011.09.004
- Levelt, W. J. M., and Wheeldon, L. R. (1994). Do speakers have access to a mental syllabary? *Cognition* 50, 239–269. doi: 10.1016/0010-0277(94)90030-2
- Lopez-Calderon, J., and Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front. Hum. Neurosci.* 8:213. doi: 10.3389/fnhum.2014.00213
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition* 25, 71–102. doi: 10.1016/0010-0277(87)90005-9
- Marslen-Wilson, W. D. (1993). “Issues of process and representation in lexical access,” in *Cognitive Models of Language Processes: The Second Sperlonga Meeting*, ed G. T. M. Altmann (Hove: Lawrence Erlbaum Associates Ltd.), 187–210.
- Mattys, S. L., White, L., and Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *J. Exp. Psychol. Gen.* 134, 477–500. doi: 10.1037/0096-3445.134.4.477
- McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0
- Mehler, J., Dommergues, J. Y., Frauenfelder, U., and Segui, J. (1981). The syllable's role in speech segmentation. *J. Verbal Learn. Verbal Behav.* 20, 298–305. doi: 10.1016/S0022-5371(81)90450-3
- Morais, J., Content, A., Cary, L., Mehler, J., and Segui, J. (1989). Syllabic segmentation and literacy. *Lang. Cogn. Process.* 4, 57–67. doi: 10.1080/01690968908406357
- Norris, D., and McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychol. Rev.* 115, 357–395. doi: 10.1037/0033-295X.115.2.357
- Obleser, J., and Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn. Sci.* 13, 14–19. doi: 10.1016/j.tics.2008.09.005
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Picton, T. W., Bentin, S. S., Berg, P. P., Donchin, E. E., Hillyard, S. A., Johnson, R., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37, 127–152. doi: 10.1111/1469-8986.3720127
- Praamstra, P., Meyer, A. S., and Levelt, W. J. M. (1994). Neurophysiological manifestations of auditory phonological processing: latency variation of a negative ERP component time-locked to phonological mismatch. *J. Cogn. Neurosci.* 6, 204–219. doi: 10.1162/jocn.1994.6.3.204
- Pylkkänen, L., and Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends Cogn. Sci.* 7, 187–189. doi: 10.1016/S1364-6613(03)00092-5
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926
- Salverda, A., Dahan, D., and McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90, 51–89. doi: 10.1016/S0010-0277(03)00139-2
- Sanders, L. D., and Neville, H. J. (2003). An ERP study of continuous speech processing. II. Segmentation, semantics, and syntax in non-native speakers. *Cogn. Brain Res.* 15, 214–227. doi: 10.1016/S0926-6410(02)00194-5
- Sanders, L. D., Newport, E. L., and Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nat. Neurosci.* 5, 700–703. doi: 10.1038/nm873
- Scharinger, M., and Felder, V. (2011). ERP signatures of cross-modal semantic fragment priming: early context effects in speech perception. *Int. J. Psychophysiol.* 80, 19–27. doi: 10.1016/j.ijpsycho.2011.01.006
- Schild, U., Röder, B., and Friedrich, C. K. (2012). Neuronal spoken word recognition: the time course of processing variation in the speech signal. *Lang. Cogn. Process.* 27, 159–183. doi: 10.1080/01690965.2010.503532
- Tyler, M., and Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *J. Acoust. Soc. Am.* 126, 367–376. doi: 10.1121/1.3129127
- Zwitserslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* 32, 25–64. doi: 10.1016/0010-0277(89)90013-9
- Zwitserslood, P. (1996). Form priming. *Lang. Cogn. Process* 11, 589–596. doi: 10.1080/016909696386980
- Zwitserslood, P. (2003). “The internal structure of words: consequences for listening and speaking,” in *Phonetics and Phonology in Language Comprehension and Production. Differences and Similarities*, eds N. O. Schiller and A. S. Meyer (Berlin: Mouton de Gruyter), 79–114. doi: 10.1515/9783110895094.79
- Zwitserslood, P. (2004). Sublexical and morphological information in speech processing. *Brain Lang.* 90, 368–377. doi: 10.1016/S0093-934X(03)00448-6
- Zwitserslood, P., Schriefers, H., Lahiri, A., and Van Donselaar, W. (1993). The role of syllables in the perception of spoken Dutch. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 260–271. doi: 10.1037/0278-7393.19.2.260

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 February 2014; accepted: 13 December 2014; published online: 12 January 2015.

Citation: Bien H, Bölte J and Zwitserslood P (2015) Do syllables play a role in German speech perception? *Behavioral and electrophysiological data from primed lexical decision.* *Front. Psychol.* 5:1544. doi: 10.3389/fpsyg.2014.01544

This article was submitted to *Language Sciences*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Bien, Bölte and Zwitserslood. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Production and perception of contrast: The case of the rise-fall contour in German

Frank Kügler\* and Anja Gollrad

Department Linguistik, Universität Potsdam, Potsdam, Germany

## OPEN ACCESS

### Edited by:

Hubert Truckenbrodt,  
Centre for General Linguistics (ZAS),  
Germany

### Reviewed by:

Laura Gonnerman,  
McGill University, Canada  
Hubert Truckenbrodt,  
Centre for General Linguistics (ZAS),  
Germany

Katrin Schweitzer,  
University of Stuttgart, Germany

### \*Correspondence:

Frank Kügler,  
Department Linguistik, Universität  
Potsdam, Karl-Liebknecht-Straße  
24–25, 14476 Potsdam, Germany  
kuegler@uni-potsdam.de

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 15 January 2014

**Accepted:** 05 August 2015

**Published:** 02 September 2015

### Citation:

Kügler F and Gollrad A (2015)  
Production and perception of  
contrast: The case of the rise-fall  
contour in German.  
Front. Psychol. 6:1254.  
doi: 10.3389/fpsyg.2015.01254

This study investigates the phonetics of German nuclear rise-fall contours in relation to contexts that trigger either a contrastive or a non-contrastive interpretation in the answer. A rise-fall contour can be conceived of a tonal sequence of L-H-L. A production study elicited target sentences in contrastive and non-contrastive contexts. The majority of cases realized showed a nuclear rise-fall contour. The acoustic analysis of these contours revealed a significant effect of contrastiveness on the height/alignment of the accent peak as a function of focus context. On the other hand, the height/alignment of the low turning point at the beginning of the rise did not show an effect of contrastiveness. In a series of semantic congruency perception tests participants judged the congruency of congruent and incongruent context-stimulus pairs based on three different sets of stimuli: (i) original data, (ii) manipulation of accent peak, and (iii) manipulation of the leading low. Listeners distinguished nuclear rise-fall contours as a function of focus context (Experiment 1 and 2), however not based on manipulations of the leading low (Experiment 3). The results suggest that the alignment and scaling of the accentual peak are sufficient to license a contrastive interpretation of a nuclear rise-fall contour, leaving the rising part as a phonetic onglide, or as a low tone that does not interact with the contrastivity of the context.

**Keywords:** production of contrast, perception of contrast, semantic-congruency task, rise-fall contour, German intonation

## 1. Introduction

This paper reports the results of a production experiment and a series of perception experiments that concern the prosodic expression of contrast in German. In particular, we investigate the phonetic details of the rise-fall contour in contexts that license either a non-contrastive or contrastive interpretation of the answer. The perception experiments seek to clarify the functional interpretation of the rise-fall contour in these contexts. In the following section a brief background on the focus-to-accent theory and the theory of intonational meaning is provided, which is mostly based on a discussion of English intonation. This discussion is followed by a brief review of German intonation and its relation to the prosodic expression of focus and contrast.

### 1.1. Focus-to-accent Theory and Intonational Meaning

Focus-to-accent theory proposes that the semantic interpretation of a focus in a sentence is distinguished from its phonological interpretation by means of the presence of a pitch accent (Gussenhoven, 1984; Selkirk, 1984). Hence, focus defined as an indication of “the presence of alternatives that are relevant for the interpretation of linguistic expressions” (Krifka, 2008, p. 247)

represents an abstract cognitive category which is prosodically expressed in language-specific ways. Syntactically, it is assumed that the focused constituent is F-marked (Jackendoff, 1972; Gussenhoven, 1984; Selkirk, 1984; Truckenbrodt, 2012). The presence of an F-mark is then assumed to have certain, language-specific effects on the phonological and phonetic expression of the focussed constituent. For instance, consider (1). While the context in (1-a) licenses the whole sentence as one of the alternatives, the context in (1-b) licenses only one particular constituent, i.e., the whale, as an alternative. The difference in F-marking in (1) then is expected to show a difference in the prosodic realization of the answer.

- (1) a. A: Erzähl mir bitte, was passiert ist. 'Please tell me, what happened?'  
 B: [ Martin hat den Wal gesehen. ]<sub>F</sub> 'Martin has seen the whale.'  
 b. A: Hat Martin den Frosch gesehen? 'Has Martin seen the frog?'  
 B: Nein. Martin hat den [ Wal ]<sub>F</sub> gesehen. 'Martin has seen the whale.'

The concept of contrast in linguistic research has a long research tradition and is generally connected to information structural categories such as topic or focus. Whether "contrast" forms its independent category in information structure (Molnár, 2002) or whether it is accompanied with either topic or focus, e.g., (Büring, 2007), remains a debate in linguistics. For an overview on this issue see (Repp, 2010). In this paper, contrast is taken in its pragmatic use for cases where it accompanies focus and corrects a given alternative from an open set of focus alternatives (cf. Krifka, 2008, for the notion of focus and corrective focus).

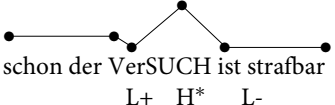
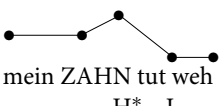
It is assumed that intonation may, depending on the language and the melody parts, carry post-lexical, sentence-level meaning (Ladd, 2008). In a compositional approach, intonational tones and their combination carry a particular meaning that a speaker may want to convey (Pierrehumbert and Hirschberg, 1990). In particular, Pierrehumbert and Hirschberg (1990) claim that the English pitch accent L+H\* carries contrastive meaning while a simple H\* pitch accents conveys the meaning of providing new information.

The effect of the two theories, focus-to-accent theory and the theory of intonational meaning, is that a pitch accent carrying a particular meaning has a preference to occur with a context that triggers this particular meaning. In other words, a L+H\* pitch accent carrying the contrastive meaning may occur more likely with a context question (1-b) that requires a contrastive interpretation of a constituent in the answer. On the other hand, speakers may produce a H\* pitch accent that carries the meaning of providing new information more likely with a context question that requires new information (1-a). Speakers may however vary their prosodic realizations since a context question may allow for different possible answers. Assume for instance that a speaker may imagine that an actual answer in (1-a) is to be contrasted with another conceivable answer. Hence, a speaker may choose, in relation to imagined additional assumptions about the context, that a contrastive contour may nevertheless be used.

## 1.2. German Intonation

The previous discussion was based on English intonation and its analysis of intonational meaning. German intonation differs from English in some respects, yet there are similar assumptions related to the meaning of H\* and L+H\* pitch accents (Grice et al., 2005). German intonation has been modeled within a number of different frameworks, e.g., in the British School approach to intonation (Klinghardt, 1925, 1927; von Essen, 1964; Pheby, 1975), in terms of different F0 peak alignments (Gartenberg and Panzlaff-Reuter, 1991; Kohler, 1991; Niebuhr, 2007), and in terms of the autosegmental-metrical approach to intonation (Uhmman, 1991; Féry, 1993; Mayer, 1997; Grabe, 1998; Barker, 2002; Braun, 2005; Gilles, 2005; Grice et al., 2005; Peters, 2005, 2006, 2014; Truckenbrodt, 2005, 2007; Baumann, 2006; Kügler, 2007; Bergmann, 2008). Related to the present discussion, work concerning F0 peak alignment has shown that the alignment of accentual peaks is related to the interpretation of information structure categories: an early peak is realized in case of given information, and a late peak in case of focused or new information (Kohler, 1991; Niebuhr, 2007).

As discussed above for English intonation (Pierrehumbert and Hirschberg, 1990), the GToBI system proposes a similar distinction in meaning between H\* and L+H\* pitch accents in German (Grice and Baumann, 2002; Grice et al., 2005). A nuclear rise-fall contour consists phonologically of a L+H\* pitch accent followed by a low phrase accent (L-), cf. (2), and the L+H\* pitch accent is assumed to carry contrastive meaning. On the other hand, a plain H\* accent is assumed to carry the meaning of newness and thus occurs preferred in non-contrastive contexts. Data on the frequency of occurrence and distribution of pitch accents in different contexts support the outlined preferences (Baumann et al., 2006; Grice et al., 2009; Sudhoff, 2010).

- (2) a.   
 schon der Versuch ist strafbar 'already the attempt is criminal'  
 L+ H\* L-  
 b.   
 mein Zahn tut weh 'my tooth is hurting'  
 H\* L-

According to Féry (1993), however, there is no phonological distinction between a pitch accent realized under contrastive and broad focus in German. Hence, the accent shapes as in (2) are analyzed with a falling H\*+L pitch accent independent of a contrastive or non-contrastive context (cf. also Grabe, 1998; Kügler et al., 2003; Peters, 2005, 2006, 2014). The varying accent shapes illustrated in (2) are all taken to constitute a rise-fall contour. Both Féry (1993) and Grabe (1998) claim that in case of a nuclear rise-fall contour the pitch rise toward the pitch peak is phonetic in nature. The assumption is that the tonal grammar of German does not exhibit a L+H\* pitch accent (Féry, 1993; Grabe, 1998).

The prosodic realization of contrast in German has been intensively studied. Generally, a focus is prosodically marked

by means of a pitch accent in German (Uhmman, 1991; Féry, 1993; Grice et al., 2009), except for particular cases of secondary focus (Féry and Ishihara, 2009; Baumann et al., 2010). Some researchers argue that slight phonetic differences between non-contrasted and contrasted realizations such as greater intensity and F0 excursion are neither necessary nor sufficient cues to signal contrast in German (Fuchs, 1976). The majority of studies however show clear and distinct differences in the production of pitch accents in non-contrasted as opposed to contrasted contexts in German (Bannert, 1985; Alter et al., 2001; Braun, 2005, 2006; Baumann et al., 2006, 2007; Féry and Kügler, 2008; Kügler, 2008; Grice et al., 2009; Sudhoff, 2010). Although the studies differ slightly in the number and kind of phonetic cues that are expressed as an interpretation of contrastiveness, generally, greater F0 excursion, or higher F0 maximum and lower F0 minimum, longer duration of the accented syllable as well as higher intensity are listed to be the relevant prosodic correlates that signal contrast in German.

This difference in prosodic marking of contrast has led to a number of studies investigating the concept of contrast in psycholinguistic research (e.g., Alter et al., 2001; Carlson, 2001; Toepel et al., 2005). As for German, the studies disagree whether or not the phonetic cues associated with contrastive accents necessarily have to be correlated with a different phonological category. Even though the prosodic cues that signal contrast in German are used to study parsing effects of ambiguous clauses, the phonological analysis of accents in contrastive contexts is still a matter of debate. Baumann et al. (2006), Braun (2006), Grice et al. (2009), and Sudhoff (2010) show that there is a considerable amount of speaker variation with respect to which pitch accent type is used in contrastive contexts compared to a neutral accentuation. In particular, Baumann et al. (2006) show that in neutral accentuation speakers tend to use downstepped accents more frequently than in case of focus, be it narrow information focus or contrastive focus (cf. also Féry and Kügler, 2008, for the preference of downstepped accents in broad focus contexts over contrastive focus contexts). However, some speakers in their study only use one identical high pitch accent independent of focus structure. The issue of speaker variation is not investigated in the current study since we are concentrating on a particular type of nuclear contour and its functional property to signal contrast.

Previous results indicate some degree of free variation with respect to accent realization, and our results of the production data show that not all speakers use raised accentual peaks in order to signal contrast. While some researchers assume a phonological difference between L+H\* and H\* pitch accents and their accompanied difference in meaning that these accents express (e.g., Grice et al., 2005; Baumann et al., 2006; Sudhoff, 2010), other researchers claim that focus and/or contrast are prosodically expressed by means of pitch register changes (e.g., Féry and Kügler, 2008; Féry and Ishihara, 2010) thus not postulating a phonological distinct representation with distinct meanings.

In order to study the prosodic expression of contrast in German, the rise-fall contour is particularly suitable since the rise can be attributed to the L+H\* pitch accent which is assumed to

carry the meaning of contrast (cf. Grice et al., 2005). On the other hand, a rise-fall contour may be realized in a broad focus context according to Féry (1993). This study will therefore examine the phonetics of the rise-fall contour in German. The contours illustrated in (2) are assumed to constitute variants of the rise-fall contour, and we used contexts that either elicited a contrastive or a non-contrastive interpretation of a particular constituent in the answer. The first question to be explored is whether speakers produce a systematic difference between rise-fall contours as a function of different contexts. The second question is whether perception tests reveal which parts of the rise-fall contour carry a functional interpretation of contrast. The next section briefly introduces methods for testing the perception of intonation.

### 1.3. Methods for Testing Intonational Categories

In intonation research a considerable body of research is concerned with the investigation of the appropriate method to test intonational categories perceptually (Gussenhoven, 1999). Different methods such as identification and discrimination studies within the categorical perception paradigm (Kohler, 1987; Gartenberg and Panzlaff-Reuter, 1991; Ladd and Morton, 1997; Remijsen and van Heuven, 1999; Post, 2000; Schneider and Lintfert, 2003; Niebuhr and Kohler, 2004; Cummins et al., 2006), imitation studies (Pierrehumbert and Steele, 1989; Redi, 2003; Dilley, 2005; Dilley and Brown, 2007; Dilley, 2010), the gating paradigm (Petrone and Niebuhr, 2014), and/or prominence judgments or semantic scales (Rietveld and Gussenhoven, 1985; Gussenhoven and Rietveld, 1988; Ladd et al., 1994) have been used and showed different success, for an overview of current methods see (Prieto, 2012).

In recent years, however, researchers emphasize the role of functional perception tests (Prieto, 2012) for the identification of tonal categories since the intonation carries function and meaning. In particular, semantic judgments were employed to test the function and meaning of intonational categories (Nash and Mulac, 1980; Gussenhoven and Rietveld, 2000; Niebuhr, 2007). Semantic congruency tests were used to study tonal categories in its appropriate context (Rathcke and Harrington, 2010; Kügler and Gollrad, 2011; del Mar Vanrell et al., 2013). The present study relies on the method of semantic congruency to test the function and meaning of the rise-fall contour in its context.

## 2. Speech Production Experiment

### 2.1. Method

#### 2.1.1. Speech Materials

The speech production experiment examines the prosodic realizations of broad and contrastive focused sentences by comparing the phonetics of the nuclear rise-fall contour in German. The experimental sentences contain the word order subject-auxiliary-object-verb (SAuxOV). The target words were embedded as objects in non-final sentence position in order to avoid any intonational phrase boundary effects. The following two factors were manipulated in order to elicit a nuclear rise-fall contour:

- The number of syllables of the target word varied between one (Wal [va:l] “whale”), two (Roman [ro.ˈma:n] “novel”), and three (Admiral [ad.mi.ˈra:l] “admiral”), all with ultima word stress. Ultima word stress was chosen to provide segmental space for a low leading tone within the accented word if such a category exists. Word-level effects on tonal alignment have been shown for English (Ladd and Schepman, 2003).
- The length of the sentence: From the basic SAuxOV structure, sentences were gradually lengthened by adding one of the two adverbials (gestern “yesterday”), and (glücklicherweise “luckily”) or a combination of both prior to the target word to increase the interaccentual distance between a prenuclear, sentence initial accent, and the nuclear accent on the target word. We expected that a larger interaccentual distance would increase the chance that speakers realize two single peak accents (Kügler et al., 2003) instead of a hat pattern, which is a frequent pattern in German (Féry, 1993; Braun, 2006). Stretching the interaccentual distance between prenuclear and nuclear accents within a sentence was not thought of being an independent factor influencing the phonetic characteristics of nuclear accents as such but was rather a strategy to ensure a large data set for the phonetic measurements. The entire material used in the production experiment is listed in the Supplementary Material.

As an experimental factor, FOCUS was manipulated eliciting broad and contrastive focus. (3-a) illustrates a context that elicits the broad focus target sentence (3-b). (4-a) illustrates a context that elicits a sentence with a contrastively focused target word (4-b). In both examples, the target word is monosyllabic.

- (3) a. Erzähl mir bitte, was passiert ist.  
'Please tell me what happened.'
- b. Maja hat den Hahn gefüttert.  
Maja has the cock fed  
'Maja has fed the cock.'
- (4) a. Hat Maja den Hund gefüttert?  
'Has Maja fed the dog?'
- b. Nein, Maja hat den Hahn gefüttert.  
No, Maja has the cock fed  
'No, Maja has fed the cock.'

The experimental sentences are highly sonorant to allow for a maximally accurate F0 analysis. Sentences were interspersed with fillers (proportion of target-filler sentences was 1 : 3) and fed into the DMDX presentation software (Forster and Forster, 2003). The experimental sentences were pseudorandomized for each subject so that sentences of the same condition did not appear adjacently and corresponding sentences had a maximal distance.

### 2.1.2. Speakers

Eight speakers participated in the experiment. All were female undergraduate students at the University of Potsdam in their twenties. All were native speakers of standard German spoken in the Berlin-Brandenburg region and reported no speech or hearing impairment. They either received course credit or were paid for participation. All subjects of this production study and

of subsequent perception experiments gave written informed consent in accordance with the Declaration of Helsinki.

### 2.1.3. Recording Procedure

For each sentence, a context eliciting broad focus (3-a) and contrastive focus (4-a), spoken by a male voice, had been previously recorded. The contexts were presented together with a target sentence both visually on screen and auditorily over headphones. The pre-recorded context sentences ensured that no uncontrolled variation of an experimenter speaking the context questions would affect the data elicitation. Speakers were asked to read and listen to the context and then to speak out the answer displayed on the screen as a response to the question. Subjects were familiarized with the task through written and verbal instructions. In case of hesitations or false starts, participants were asked to repeat the sentence. Recordings took place in a sound-proof chamber equipped with an AT4033a audiotechnica studio microphone, using a C-Media Wave sound card at a sampling rate of 44.1 kHz with 16 bit resolution. Presentation flow was controlled by the experimenter, and participants were allowed to take a break at any point. A total of 384 target sentences (8 speakers × 2 focus conditions × 6 target words × 4 sentence lengths) had been recorded.

### 2.1.4. Grouping of Nuclear Contours

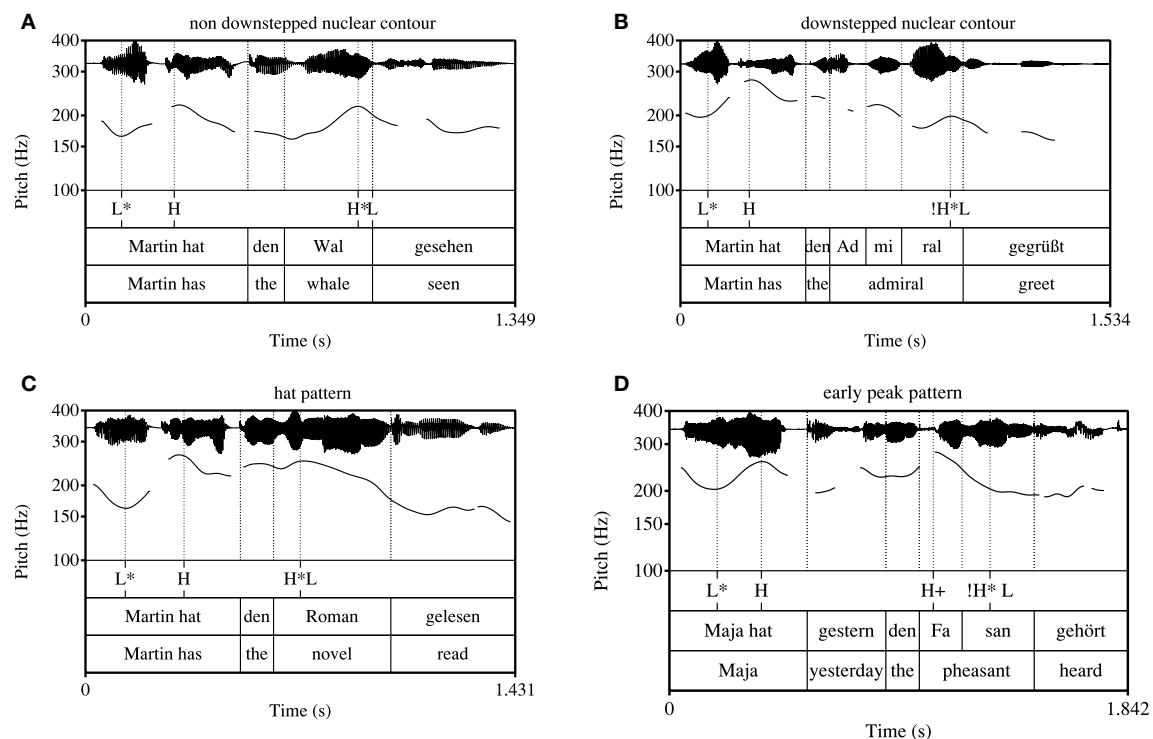
As there is a range of possible nuclear intonation contours in German (Féry, 1993; Grabe, 1998; Grice et al., 2005, 2009), we grouped nuclear contours according to their overall shape. Since we are interested in the nuclear rise-fall contour, we separated these from other nuclear contours. We established four different nuclear contours in our data which are illustrated in **Figure 1**. The annotation of the pitch contours was based on the tonal grammar of German proposed by Féry (1993). The total of 384 sentences was subgrouped into the four distinct phonological contours as follows:

Subgroup (a) contains 255 non-downstepped nuclear rise-fall contours, which comprise contours that contain either a prenuclear rising or falling accent (cf. Uhmann, 1991; Féry, 1993). **Figure 1A** illustrates a nuclear rise-fall contour with a prenuclear rising accent. The two accents in **Figure 1A** have comparable F0 scaling concerning their H tones. The instances of the rise-fall contour constitute the cases for further phonetic analysis.

Subgroup (b) contains 25 downstepped nuclear rise-fall contours, which comprise either rising or falling prenuclear accents. The H tone of the nuclear accent in **Figure 1B** is scaled lower relative to its preceding H tone of the prenuclear accent, which causes the perceptual impression of a downstep. The downstep is indicated by the exclamation mark. Note that downstepped accents lack clear low turning points in F0 prior to the downstepped peak in most of the cases. Nuclear downstepped accents are used frequently in German (Féry and Kügler, 2008; Grice et al., 2009).

The third subgroup (c) consist of 36 hat patters (Kohler, 1991; Uhmann, 1991; Féry, 1993; Braun, 2006) (cf. “bridge accent” in Wunderlich, 1988), which results from a prenuclear rising or high pitch accent and a nuclear falling pitch accent. Both accents are





**FIGURE 1 | Pitch track representation of the four phonological contours, from top left to bottom right: (A) non downstepped nuclear contour, (B) downstepped nuclear contour, (C) hat pattern, and (D) early peak pattern.** All four contours involve a prenuclear accent followed by a nuclear accent. The target word carries the nuclear accent, which is associated with the most prominent syllable.

concatenated by a high F0 plateau, without a dip between the prenuclear and the nuclear accent (cf. **Figure 1C**).

The fourth subgroup (d) contains 68 other types of nuclear accents, such as early peaks. This category displays cases of a prenuclear accent followed by a nuclear accent, where the nuclear accent displays a different alignment shape as the ones before. In **Figure 1D** the peak of the falling accent is aligned with the syllable preceding the stressed syllable of the target word, a case referred to as early peak (Kohler, 1991; Uhmman, 1991; Féry, 1993; Grice et al., 2005).

Both authors conducted the grouping independently and agreed in about 92% of the cases. For the remaining cases, we discussed each individual contour by listening and looking at the F0 contour to eventually decide on the contour.

As can be seen in **Table 1**, the non-downstepped rise-fall contours are almost equally distributed across the two context conditions. In 45% of the 255 cases, a rise-fall was realized in a broad focus context, somewhat more (55%) in a context eliciting contrastive focus. In these realizations we analyzed how a contrastive focus changes the phonetic realization of the rise-fall. **Table 1** also shows that 19 downstepped accents and 29 hat patterns are preferred realizations in a broad focus context (80 and 76%, respectively), which is in line with previous findings (cf. Grice et al., 2009). In the following, group (a) is investigated in more detail.

### 2.1.5. Data Processing

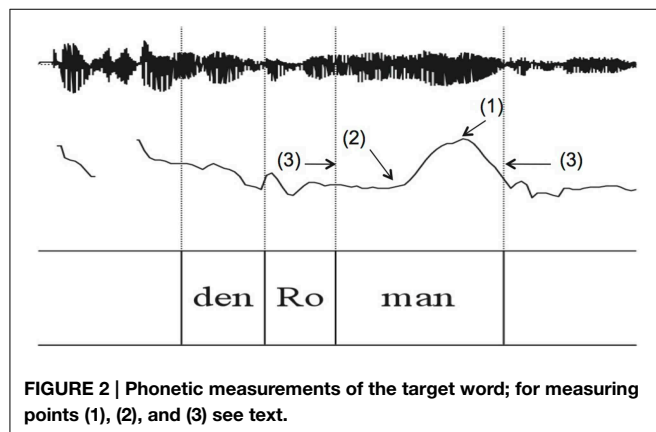
The 255 experimental sentences of group (a) were hand-annotated and subjected to phonetic analysis using Praat software (Boersma and Weenink, 2013). The annotation comprised the target noun phrase including the determiner, see **Figure 2**. Annotation was done on the level of the syllable. The following phonetic measurements were conducted, numbers correspond to measuring points in **Figure 2**:

- The pitch peak (H) of the target words in Hertz (Hz), see point (1) in **Figure 2**
- The corresponding time of the peak ( $t_H$ ), see point (1) in **Figure 2**
- A low turning point in pitch prior to the peak (l) in Hz, which corresponds to the “elbow” measure in D’Imperio (2000), see point (2) in **Figure 2**
- The corresponding time of the low turning point ( $t_l$ ), see point (2) in **Figure 2**
- The beginning and the end of the accented syllable ( $t_{beg}$ ,  $t_{end}$ ), see point (3) in **Figure 2**

Pitch analysis was conducted using a Hanning window of 0.4 s length with a default 10 ms analysis frame. The pitch contour was smoothed using the Praat smoothing algorithm (frequency band 10 Hz) to diminish microprosodic perturbations. Out of these phonetic measurements, the following variables were calculated:

**TABLE 1 | Distribution of nuclear contours per subgroup, split by focus condition.**

Phonological contour	Broad focus	Contrastive focus	n
(a) Non-downstepped	115	140	255
(b) Downstepped	19	6	25
(c) Hat pattern	29	7	36
(d) Other (early peak)	29	39	68
Sum	192	192	384

**FIGURE 2 |** Phonetic measurements of the target word; for measuring points (1), (2), and (3) see text.

- The excursion (E) between the low turning point and the peak:  $E[\text{Hz}] = H - l$ .
- The velocity (V) of the pitch rise:  $V[\text{Hz/s}] = \frac{H-l}{t_H-t_l}$ .
- The relative alignment of the the pitch peak (A – H) with reference to the end of the accented syllable divided by the accented syllable's duration:  $A_H[\%] = \frac{t_H-t_{\text{end}}}{t_{\text{end}}-t_{\text{beg}}} * 100$ .  
The end of the accented syllable was chosen based on the results of Grabe (1998) who showed alignment of H\* tones at the right edge of the accented syllable's rime.
- The relative alignment of the low turning point (A – l) with reference to the beginning of the accented syllable divided by the accented syllable's duration:  $A_l[\%] = \frac{t_l-t_{\text{beg}}}{t_{\text{end}}-t_{\text{beg}}} * 100$ .
- The duration (D) of the accented syllable:  $D[\text{ms}] = t_{\text{end}} - t_{\text{beg}}$ .

### 2.1.6. Statistical Analysis

The results of the phonetic calculation were evaluated against the fixed factor FOCUS [with the two levels broad focus (BF) and contrastive focus (CF)] using linear mixed models (Bates et al., 2013). The reference level in the models was BF. The models applied crossed random factors speaker and item. Random slopes (Barr et al., 2013) for speakers and items were integrated into the models assuming that differences exist for each speaker's individual pitch range. Backward modeling (Barr et al., 2013) of random slopes for speaker and item was applied, and likelihood ratio tests were run to evaluate the models. The basis for removing factors was a *p*-value of the likelihood ratio test of  $p < 0.05$  and lower AIC values.

**TABLE 2 | Report of the linear mixed-effects models for each of the measured cues.**

Variable		Coefficients	SE	t-value	Sign.
L (Hz)	Intercept	191.187	6.623	28.439	
	CF	−2.579	2.271	−1.136	n.s. <sup>a</sup>
H (Hz)	Intercept	233.67	8.415	27.769	
	CF	4.728	2.471	1.913	(*) <sup>a</sup>
E(Hz)	Intercept	42.416	6.287	6.746	
	CF	7.412	3.524	2.103	* <sup>a</sup>
V (Hz/s)	Intercept	251.19	31.52	7.969	
	CF	46.85	11.03	4.247	* <sup>b</sup>
A–L (%)	Intercept	14.275	4.568	3.125	
	CF	4.940	4.401	1.122	n.s. <sup>a</sup>
A–H (%)	Intercept	19.98	1.671	11.953	
	CF	−8.31	2.223	−3.738	* <sup>a</sup>
D (ms)	Intercept	243.297	8.361	29.10	
	CF	15.581	3.898	3.997	* <sup>b</sup>

<sup>a</sup> Based on a linear mixed model including item with random intercepts and subjects with random intercepts and random slopes.

<sup>b</sup> Based on a linear mixed model including item and speakers with random intercepts only.

\* indicates significance at the level  $p < 0.05$ ; n.s. refers to non-significance.

## 2.2. Results

The statistical results are shown in Table 2. For each individual variable it is shown which model presents the best fit. Significance at the level  $p < 0.05$  for a factor was determined with an absolute *t*-value of 2 or greater (Barr et al., 2013). We find a significantly lower excursion size (E) in BF compared to CF (means for BF: 45.7 Hz and CF: 52.2 Hz), a significantly slower velocity of the rise (V) in BF compared to CF (means for BF: 262.6 Hz/s vs. CF: 305.0 Hz/s), a significantly earlier alignment of the accentual peak in relation to the end of the syllable (A–H) in BF compared to CF (means for BF: 20.30% and CF: 11.78%), and a significantly shorter duration (D) in BF compared to CF (means for BF: 247 ms and CF: 261 ms). The model for the scaling of the accentual peak (H) reveals a near significant effect between both focus conditions (means for BF: 236.1 Hz and CF: 240.8 Hz). The analysis reveals that in contrastive contexts the accentual peak is affected. It is realized higher and it occurs later. In absolute values, the low turning point prior to the accentual peak [L (Hz)] does not differ systematically between both focus conditions (means for BF: 190.4 Hz and CF: 188.6 Hz), nor does the relative alignment of the low turning point (A–L) differ between focus conditions (means for BF: 12.39% and CF: 18.94%). The fact that the velocity of the rise and the excursion size show a significant effect is compatible with the change being located only in the H.

## 2.3. Discussion

The analysis of the phonetic variables yields no clear indication that the low F0 turning point prior to the accentual peak represents a systematic difference between the two focus contexts. Neither the model for L-tone scaling nor the model for L-tone alignment showed a systematic difference between a broad and a contrastive context. On the other hand, the model for scaling and the model for alignment of the accentual peak showed differences

as a function of focus context. The scaling of the accentual H tone is higher in contrastive focus contexts, which is well in line with previous findings (Bannert, 1985; Alter et al., 2001; Braun, 2005, 2006; Baumann et al., 2006, 2007; Féry and Kügler, 2008; Grice et al., 2009; Sudhoff, 2010); the effect approaches significance. The significantly increased duration is also well-known for German (cf. e.g., Kügler, 2008).

The fact that H-tone scaling only approaches significance seems to be due to the fact that not all speakers employ this strategy to realize contrastive focus. Model comparison for H-tone scaling applying likelihood ratio tests revealed that when removing the slope factor for the random effect of speaker, the effect of FOCUS on the height of the H becomes significant ( $coef = 2.626, SE = 0.841, t = 3.122$ ). Thus, the best fit model in **Table 2** including the slope effect for speakers indicates speaker-specific differences. We also calculated the individual speaker means which showed that speakers differed considerably in their scaling of the H-tone. Given this finding plus the fact that the model for alignment of the H-tone additionally showed that all speakers employed significantly later accentual peaks in case of contrastive focus contexts suggests that only some speakers employ a different scaling as a means to express contrastive focus. Individual speaker strategies in prosodic focus marking have been reported earlier for German (Baumann et al., 2006). In addition, perception tests showed that the strategy of a higher or later peak revealed identical effects of signaling increased prosodic prominence (Ladd and Morton, 1997). Additionally, duration serves as a robust cue to signal prosodic prominence, and we also found a systematic increase in duration in contrastive contexts.

Furthermore, the phonetic effects triggered by focus should be seen in relation to prenuclear accents. The utterances realized under broad focus exhibit a F0-lowering from prenuclear to nuclear accents, while it is the other way around for the utterances realized under contrastive focus (see **Table 3**). Paired-samples *t*-tests for broad focus and contrastive focus show that the scaling of prenuclear and nuclear high tones differs significantly. Similar patterns of a relational scaling of pitch accents are reported in Féry and Kügler (2008).

Taken together, the results of the production study indicate that speakers realize a phonetic difference in intonation as a function of the focus condition. In the following series of studies we will test which parts of the rise-fall contour interact perceptually with the contrastivity of the context.

**TABLE 3 | Mean F0 maximum in Hz of the prenuclear and nuclear accents, split by focus condition.**

Focus condition	Prenuclear accent	Nuclear accent	Welch two sample <i>t</i> -test
Broad focus	263.1	236.1	$t = 5.7551$ , $df = 188.384$ , $p < 0.001$
Contrastive focus	234.0	240.8	$t = -2.0813$ , $df = 269.135$ , $p < 0.05$

### 3. Speech Perception Experiments

A series of semantic congruency tasks investigate whether German listeners use the phonetic differences shown in the production study to distinguish the rise-fall contour between contexts that elicit broad or contrastive focus. Semantic congruency tests have been successfully used to explore the perception of functional intonation contrasts (Rathcke and Harrington, 2010; Kügler and Gollrad, 2011; Prieto, 2012; del Mar Vanrell et al., 2013). The test allows us to evaluate the degree of perceived appropriateness of target intonation patterns within different pragmatic contexts.

#### 3.1. Perception Experiment 1: Original Data

##### 3.1.1. Material

The first experiment investigates whether the acoustic differences found in the production data are perceived as an indicator for the appropriate context they were realized in. Following different perception studies that rely on the speech of one speaker (Kohler, 1991; Niebuhr, 2007; Dille and Heffner, 2013) stimulus materials were taken from one of the speakers of the production study. To choose from the eight speakers of the production study, we decided to choose a speaker who produced the most prominent difference from the mean value of the low turning point in both focus conditions.

The target sentences correspond to the six SAuxOV sentences from the production study (cf. Supplementary Material). Each one was uttered in broad focus (BF) and contrastive focus contexts (CF) resulting in 12 sentences. The semantic congruency experiment consisted of these 12 target sentences where intonation was congruent with the pragmatic context (6 BF–BF dialogs, 6 CF–CF dialogs), and 12 cross-spliced target sentences where intonation was incongruent with the pragmatic context (6 CF–BF dialogs, 6 BF–CF dialogs). Stimuli were scaled at an intensity of 70 db. Each dialog was presented 3 times which resulted in a total of 72 dialogs per experiment. The stimuli were auditorily presented over headphones with the MFC Praat software (Boersma and Weenink, 2013). Participants were asked to listen to each dialog carefully and then evaluate whether they regard the intonation of the target sentence to the given context as “congruent” or as “incongruent” (by clicking either on the “congruent box” or the “incongruent box” visible on the screen). After written and verbal instructions, a test run of 3 dialogs was carried out before the experiment started. The experiment lasted approximately 20 min.

##### 3.1.2. Participants

Thirty-six participants took part in the experiment (10 male, 26 female). They were all undergraduates in their twenties, reported no speech or hearing deficits, and were naïve with respect to the purpose of the study. They were either paid for participation or received course credits.

##### 3.1.3. Hypothesis

For the factor CONGRUENCY we hypothesize that congruent dialogs (BF–BF and CF–CF pairs) are rated more congruent than incongruent dialogs (CF–BF and BF–CF pairs). This hypothesis reflects the fact that the stimuli produced in their original (=

congruent) context are assumed to be perceived as fitting well with their context while cross-spliced context-answer stimuli should create an incongruent impression. As for the factor CONTEXT we assume no particular effect. In other words, both broad and contrastive contexts are assumed to create the same amount of variation in the perceptual impression.

### 3.1.4. Results

**Figure 3** displays the rate of congruent responses in percentage to all dialog types, separated into BF-context (left bars) and CF-context (right bars). In general, the appropriateness of the target intonation pattern to a context was rated higher for congruent (BF–BF and CF–CF) than for incongruent dialog types (CF–BF and BF–CF). Specifically, in 61.9% of the BF–BF dialogs, and in 79.3% of the CF–CF dialogs, the target intonation was rated as congruent to its context, while for incongruent dialogs the number of congruent responses was reduced to 47.2% in BF–CF dialogs, and to 59.4% in CF–BF dialogs.

For the statistical, frequency-based analysis, we fit a multilevel model (Bates et al., 2013) using crossed random factors participant and item applying random intercepts and slopes, and CONTEXT (with levels BF/CF) and CONGRUENCY (with levels congruent/incongruent) as fixed factors. The analysis relied on the choice of answer (congruent vs. incongruent) as a dependent variable. Treatment-coding was applied using level *BF* of the factor CONTEXT as baseline, and level *incongruent* of the factor CONGRUENCY as baseline. Model comparison for the random effect structure was applied, which was based on the same method as described in Section 2.1.6 above.

The model representing the best fit used both random slopes and intercepts of speaker and item for both fixed factors. The model reveals a significant effect for CONGRUENCY, but neither for CONTEXT nor for the interaction, cf. **Table 4**.

### 3.1.5. Discussion

The semantic congruency task revealed that listeners judged congruent dialogs as more congruent than incongruent dialogs. The expected effect of CONGRUENCY was thus borne out. Listeners rely on the phonetic cues in the nuclear rise-fall contour

that signal contrastive or non-contrastive interpretations. This result also shows that listeners are able to perceive the subtle acoustic differences that were produced in different contexts. This allows us to continue to investigate which of the acoustic cues, i.e., the accentual high tone or a low turning point in F<sub>0</sub>, are necessary to perceive the functional difference.

Two subsequent perception experiments were carried out to determine whether the phonetic difference of the high peak or of the low turning point is functionally relevant. The high peak and the low turning point were manipulated separately from each other in two different experiments. The next section describes the phonetic manipulation of the accentual peak on listeners' interpretation in relation to contrast, the third perception experiment investigates the role of the low turning point itself.

## 3.2. Perception Experiment 2: Manipulation of the H\* Accent

Given that original stimuli are appropriately categorized according to focus contexts (perception Experiment 1), and in line with previous findings on the effect of contrast on accentual peaks (Ladd and Morton, 1997; Gussenhoven, 2004; Baumann et al., 2006; Féry and Kügler, 2008), we predict that the F<sub>0</sub> peak height is functionally relevant, i.e., a higher F<sub>0</sub> peak is expected to cause a perceptual impression of contrast.

### 3.2.1. Speech Material

We test this prediction by manipulating the scaling of the H\* accent successively. The sentences for the H\* manipulation were taken from the same speaker used for the first experiment. To keep the total amount of stimuli in a manageable size for a perception study, a total of four target sentences including disyllabic and trisyllabic target words were chosen for the manipulation procedure. These sentences were realized in broad focus contexts, and in contrastive focus contexts yielding eight sentences in total. For each of the 4 sentences, the manipulation of the H\* peak was done in relation to the corresponding prenuclear accent on the subject; **Figure 4** illustrates this relationship between prenuclear and manipulated nuclear accents. For each sentence, the maximum F<sub>0</sub> value on the prenuclear accent was calculated. By adding 50 Hz and by subtracting 30 Hz from the calculated F<sub>0</sub> maximum of the prenuclear accent, we defined the manipulation range separately for each sentence. This range corresponds roughly to two standard deviations from the mean F<sub>0</sub> value of the nuclear accent peak gained from the production data. The H\* accent was manipulated with a Praat script, such that for each original sentence, five stimuli with varying values for the H\* peak were re-synthesized; **Figure 4** illustrates a horizontal line from the prenuclear peak to the nuclear accent showing two stimuli with lower nuclear peaks, two stimuli with higher nuclear peaks, and one stimulus with identical pitch height compared to the prenuclear accent. Each manipulated target sentence was concatenated with an originally congruent context question (BF–BF, CF–CF) and with an originally incongruent context question (CF–BF, BF–CF), resulting in a total of 80 stimuli (4 sentences × 2 focus conditions × 2 contexts × 5 manipulations). All stimuli

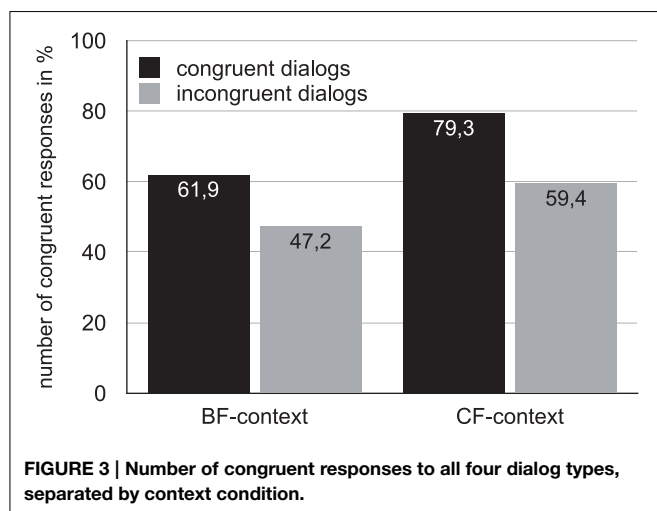
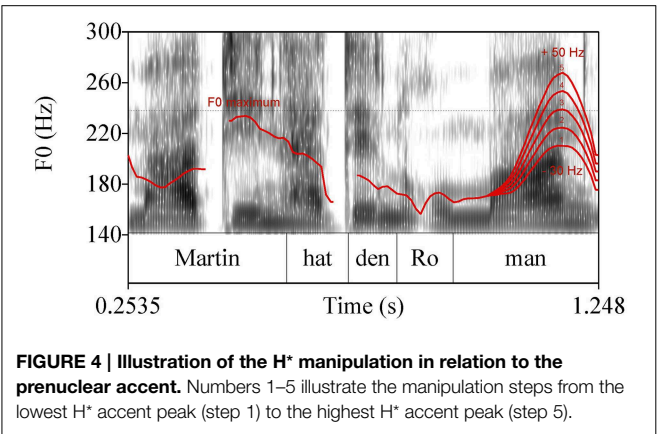




TABLE 4 | Report of the linear mixed effects model specified in the text with congruent/incongruent ratings as dependent variable.

	Coefficients	SE	z-value	Sign.	Pr(>  z )
(Intercept)	0.6673	0.1297	5.143	*	0.00001
context = CF	0.8405	0.5234	1.606	n.s.	0.10833
congruency = congruent	0.9649	0.2780	3.470	*	0.00052
Interaction	0.2596	0.2651	0.979	n.s.	0.32735

\* indicates significance at level  $p < 0.05$ , n.s. refers to non-significance.



were scaled at an intensity of 70 db. Stimuli were subdivided into two lists of 40 stimuli each, such that a participant would hear the same target word originally spoken in one focus condition once in the original matching context and once cross-spliced in the non-matching context. Each list contained 20 congruent and 20 incongruent dialog pairs. The precise grouping arrangement is listed in the Supplementary Material. The reason to divide the stimuli into two lists was to present listeners a comfortable number of dialogs to be evaluated. The experimental task was identical to the one of perception Experiment 1, except that the 80 stimuli were divided into two sets. Participants listened to either set 1 or set 2. The experiment lasted approximately 15 min.

3.2.2. Participants

Forty-eight undergraduate students from Potsdam University (13 male, 35 female) participated in the experiment. They were native speakers of German in their twenties and reported no speech or hearing impairment. The participants were naïve as to the purpose of the experiment and did not participate in perception Experiment 1. Each participant received course credit for participation. Participants were divided into two groups to listen to either the first or the second experimental set.

3.2.3. Hypothesis

If a phonetic cue for contrastiveness (e.g., a higher H\*), has an effect on the perception of contrast, it will influence the congruency ratings in the two contexts differently: In a contrastive context condition, an effective cue for contrastiveness will lead to more congruency judgements. In a non-contrastive context, an effective cue for contrastiveness will lead to less congruency judgements. For the H\* accent manipulation, we

expect thus that for contrastive contexts higher F0 peaks (manipulation step 5) cause a perceptual impression of contrastiveness, both in originally congruent (CF-CF) and originally incongruent dialogs (CF-BF) (cf. Baumann et al., 2007; Féry and Kügler, 2008 for higher F0 peaks in German). For broad focus contexts, we expect that lower F0 peaks (manipulation step 1) cause a perceptual impression of broad focus, both in originally congruent (BF-BF) and originally incongruent dialogs (BF-CF); lower peaks are assumed to correspond to the downstep pattern in German broad focus sentences (Féry and Kügler, 2008; Grice et al., 2009). Therefore, we predict that an effective cue for contrastiveness will show an interaction of MANIPULATION and CONTEXT on the dependent variable congruency.

3.2.4. Results

Figure 5 displays the results of the H\* manipulation experiment separated for the highest H\* accent manipulation step 5 (left-hand bars) and the lowest H\* accent manipulation step 1 (right-hand bars) for each dialog type. In all contrastive context dialogs under manipulation step 5 (CF-CF and CF-BF), a higher H\* accent of the target word leads to a higher number of congruency ratings (both 78.1%) compared to the corresponding dialogs under manipulation step 1 (between 39.6 and 51%). In all broad focus context dialogs under manipulation step 5 (BF-BF and BF-CF), a higher H\* accent of the target word leads to approximately identical congruency ratings compared to the corresponding dialogs under manipulation step 1 (ranging between 61.4 and 72.9%).

As described for perception Experiment 1, Section 3.1.4, we fit a multilevel model with CONTEXT (with levels BF/CF) and MANIPULATION (with levels step1/step5) as fixed factors, and calculated likelihood ratio tests on the basis of backward modeling of the random factors to identify the best fit model. Note that only a subset of the data entered into the analysis, i.e., ratings for the endpoints of the manipulation range, step1 and step5, respectively. This was done to evaluate an effect of the maximal manipulation on the perception; an analysis of the step-wise manipulation is given below. Treatment-coding was applied using level BF of factor CONTEXT, and level step1 of factor MANIPULATION as baseline. The best fit model used random intercepts and slopes of both fixed factors for subjects, and neither random slopes nor intercepts for item. The model reveals a significant interaction of MANIPULATION and CONTEXT, as well as a significant effect for MANIPULATION, but no effect for CONTEXT alone, cf. Table 5. According to the hypothesis, a higher H\* accent realization is an effective cue for contrastiveness



due to the significant interaction. Higher F0 peaks were expected to be congruent in contrastive contexts, and lower F0 peaks in broad focus contexts independent of stimulus origin.

We computed a Pearson product-moment correlation coefficient to assess the relationship between the manipulation steps and the congruency ratings, separately for each dialog type. **Figures 6C,D** show that in the contrastive focus context dialogs, a positive correlation between dialog type and manipulation step is evident: an increasing value of the H\* peak (manipulation step 1 = low H\* value; manipulation step 5 = high H\* value), raises the number of congruent responses (CF–CF:  $r = 0.298$ , CF–BF:  $r = 0.204$ ). On the other hand, in all broad focus context dialogs (**Figures 6A,B**), the H\* peak manipulation does not influence the rating. There was a close to zero correlation between manipulation step and congruency rating (BF–BF:  $r = -0.08$ , BF–CF:  $r = -0.016$ ). Congruent responses remain at an equal high level, independently of the height of the H\* accent.

### 3.2.5. Discussion

The results reveal two major aspects. First, the manipulation of the pitch peak has a significant effect on the interpretation of the pitch accent. The higher the peak the more often were stimuli rated as congruent in the contrastive focus context. This result was independent of stimulus origin, i.e., whether a stimulus was originally uttered in a broad or contrastive context did not affect its interpretation. It is thus the F0 height (in relation to the previous pitch accents) that caused the perception of contrastiveness in the experiment. This result is in line with previous findings and assumptions on the relationship

between contrastive focus and its prosodic realization in German (Bannert, 1985; Alter et al., 2001; Braun, 2005, 2006; Baumann et al., 2006, 2007; Féry and Kügler, 2008; Grice et al., 2009; Sudhoff, 2010).

Second, the obtained significant effect for MANIPULATION points to the fact that the two contexts allow a different amount of prosodic variation. In contrastive contexts, it was clearly the peak manipulation that mattered, and hence, only a certain amount of variation regarding pitch peak scaling was tolerated by listeners. In broad focus contexts, however, listeners accepted both, lower and higher F0 peaks as congruent prosodic realizations, again, independent of stimulus origin. This perceptual behavior mirrors the free variation found in the production of German broad focus contours: Féry and Kügler (2008) showed that downstepped and upstepped pitch accents occur equally frequent (45.7–54.3%) in broad focus contexts. Downstep and upstep correspond in our experiment to the manipulation of the pitch peak, lower scaling refers to downstep, higher scaling to upstep in relation to the prenuclear accent (cf. **Figure 4**).

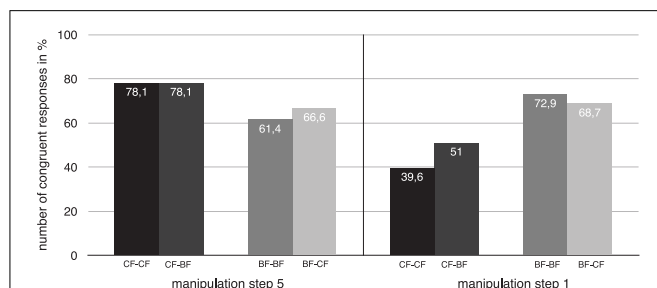
Given the significant interaction of MANIPULATION and CONTEXT we can conclude that the higher scaling of the H\* accent reflects a perceptual interpretation of contrastiveness. The next experiment examines whether a manipulation of the low turning point prior to the H\* peak can be attributed to a perceptual interpretation of contrastiveness as well, as postulated by Grice et al. (2005).

## 3.3. Perception Experiment 3: Manipulation of the Low Turning Point

### 3.3.1. Material

This experiment investigates the role of the low turning point in F0 of the nuclear rise-fall contour, more specifically the issue whether the height of the low turning point interacts with the contrastivity of the context. The sentences for the low turning point manipulation were the same as the ones used for perception Experiment 2. Each sentence was manipulated at the position of the low turning point, cf. **Figure 7**.

Using a Praat script, manipulation procedure was as follows: The F0 contour of the original file was stylized. The F0 points at the onset of the target word and at the accentual peak were retained and the F0 points between them were deleted. At the time of the label of the low turning point (see production study) a pitch point was inserted, and pitch was interpolated between the remaining pitch points. The end points of the F0 height continuum of the inserted pitch points were determined relative

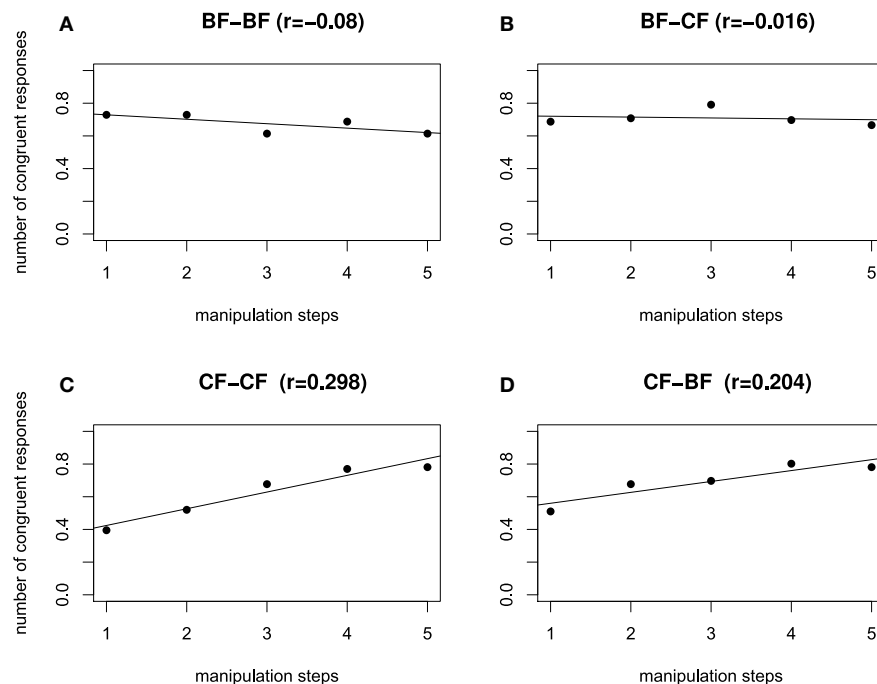


**FIGURE 5 |** Number of congruent responses to all dialog types, separated by manipulation step 5–highest H\* peak (left-hand bars) and manipulation step 1–lowest H\* peak (right-hand bars); black and dark gray bars indicate dialog pairs with contrastive contexts, lighter gray bars indicate dialog pairs with broad focus contexts.

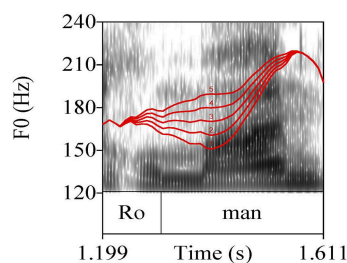
**TABLE 5 |** Report of the linear mixed effects model with the fixed factors context and manipulation and with congruent/incongruent ratings as dependent variable.

	Coefficients	SE	z-value	Sign.	Pr(>  z )
(Intercept)	0.8907	0.1625	5.482	*	0.00001
context = CF	−0.3338	0.3401	−0.982	n.s.	0.32629
manipulation = step5	0.9754	0.2031	4.804	*	0.00001
Interaction	0.9477	0.2747	3.450	*	0.00056

\* indicates significance at level  $p < 0.05$ , n.s. refers to non significance.



**FIGURE 6 | Influence of H\* manipulation on the number of congruent responses, separated by dialog type, for all manipulation steps, starting from left to right with step 1 = low H\* to 5 = high H\*. Congruent dialog types BF-BF (A) and CF-CF (C), incongruent dialog types BF-CF (B) and CF-BF (D).**



**FIGURE 7 | Illustration of the low turning point manipulation.** Numbers 1–5 illustrate the manipulation steps from the lowest turning point (step 1) to the highest turning point (step 5).

to the F0 height that was produced in the utterance. A distance of two standard deviations from the mean in both directions resulted in a manipulation range from 150 to 190 Hz for each sentence. Thus, five stimuli with a difference of 10 Hz between the low turning points were created, cf. **Figure 7**.

Each manipulated target sentence was concatenated with an originally congruent context question (BF–BF, CF–CF) and with an originally incongruent context question (CF–BF, BF–CF), resulting in a total of 80 target sentences (4 sentences  $\times$  2 focus conditions  $\times$  2 contexts  $\times$  5 manipulations). These 80 target sentences were scaled at an intensity of 70 db, and stimuli were subdivided into two lists of 40 stimuli each (see the Supplementary Material for the stimuli and their groupings). The experimental task was identical to that one of perception Experiment 2. The experiment lasted approximately 15 min.

### 3.3.2. Participants

Forty-eight undergraduate students from Potsdam University (16 male, 32 female) with no hearing deficits took part in this perception experiment. They did not take part in the first or second perception experiment. They were all in their twenties, and were either paid for participation or received course credit points. Participants were divided into two groups to listen to either the first or the second experimental set.

### 3.3.3. Hypothesis

As in the previous experiment, we predict a significant interaction of the factors MANIPULATION and CONTEXT based on the assumption that the low turning point in F0 interacts with the contrastivity of the context. We expect a lower F0 turning point to signal contrast, cf. the difference of the schematic contours in (2). The prediction thus is that independent of stimulus origin (originally uttered in a broad or a contrastive context), lower F0 turning points should cause significantly more congruent answers in contrastive contexts. Similarly, higher F0 turning points prior to the accentual peak should cause significantly more congruent answers in broad focus contexts.

### 3.3.4. Results

**Figure 8** depicts the number of congruent responses to all dialog types, separated for the highest low turning point manipulation step 5 (left-hand bars) and the lowest low turning point manipulation step 1 (right-hand bars). Independent of manipulation step, congruent context-target dialogs (CF–CF, BF–BF) obtained an equal high number of congruency ratings, while

incongruent context-target dialogs (CF–BF, BF–CF) obtained an equal low number of congruency ratings.

As described for perception Experiment 1, Section 3.1.4, we fit a multilevel model with CONTEXT (with levels BF/CF) and MANIPULATION (with levels step1/step5, i.e., the endpoints of the manipulation range) as fixed factors, and calculated likelihood ratio tests on the basis of backward modeling of the random factors to identify the best fit model. As before, only the endpoints of the manipulation range entered the analysis. Treatment-coding was applied using level *BF* of factor CONTEXT, and level *step1* of factor MANIPULATION as baseline. The best fit model used crossed random factors participant and item, applying random slopes and intercepts for both fixed factors with participants, and random slopes with item for the fixed factor CONTEXT. The model reveals no significant interaction, and no significant effect of the fixed factors CONTEXT and MANIPULATION, cf. **Table 6**. According to our hypothesis, the factor MANIPULATION was defined such that the lowest manipulation step should result in a contrastive interpretation. Thus, the lowest manipulation step was expected to be rated more congruent in contexts that require a contrastive interpretation in the answer. Consequently, the highest manipulation step should result in a non-contrastive interpretation, thus should be rated more congruent in contexts that require a non-contrastive interpretation of the answer.

As for Experiment 2, we computed a Pearson product-moment correlation coefficient to assess the relationship between the manipulation steps and the congruency ratings, separately

for each dialog type. **Figure 9** shows no correlation between manipulation step and congruency ratings for either of the dialog pairs. In other words, the close to zero correlations show that the manipulation had no influence on the congruency rating, which is in line with the non-significant interaction of the factors CONTEXT and MANIPULATION, cf. **Table 6**. However, **Figure 9** shows a difference in level of congruency ratings, i.e., congruent dialogs were rated more congruent (cf. **Figures 9A,C**) than incongruent ones (cf. **Figures 9B,D**).

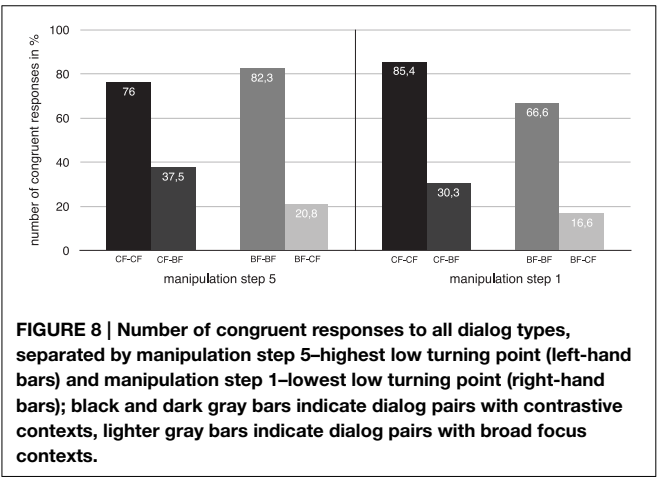
3.3.5. Discussion

The results of the manipulation of the low F0 turning point reveal two aspects. First, independently of the prosodic manipulation, congruent context-target dialogs were rated better than incongruent dialogs. Second, the non-significant interaction of MANIPULATION and CONTEXT suggest that the low turning point before the accentual peak does not contribute to the perceptive impression of contrast. If it would, it was expected that the number of congruency ratings for manipulations CF–CF:5 (190 Hz) and BF–BF:1 (150 Hz) would have been considerably lower, likewise the number of congruency ratings for manipulations BF–CF:1 (150 Hz) and CF–BF:5 (190 Hz) would have been higher. Taken the results of the H\* manipulation from the previous experiment together with the results of this experiment suggest that the higher scaling of the H\* accent stimuli is the relevant cue that signals contrastivity perceptually in German.

4. Discussion and Conclusion

This study was concerned with the phonetics of the nuclear rise-fall contour in German. In particular, we investigated how the phonetic realization of the rise-fall contour interacts with contexts that require a contrastive or broad focus interpretation in the answer. To this end, a production experiment and a series of perception experiments were carried out. The analysis of the production data revealed that contrastive focus changes the phonetics of the rise-fall contour. Speakers realized significantly higher and later F0 peaks in contrastive contexts. The realization of the low turning point prior to the accentual peak showed no significant differences. The fact that contrastive focus raises nuclear H\* accents in German confirms earlier results (Baumann et al., 2006, 2007; Féry and Kügler, 2008; Grice et al., 2009).

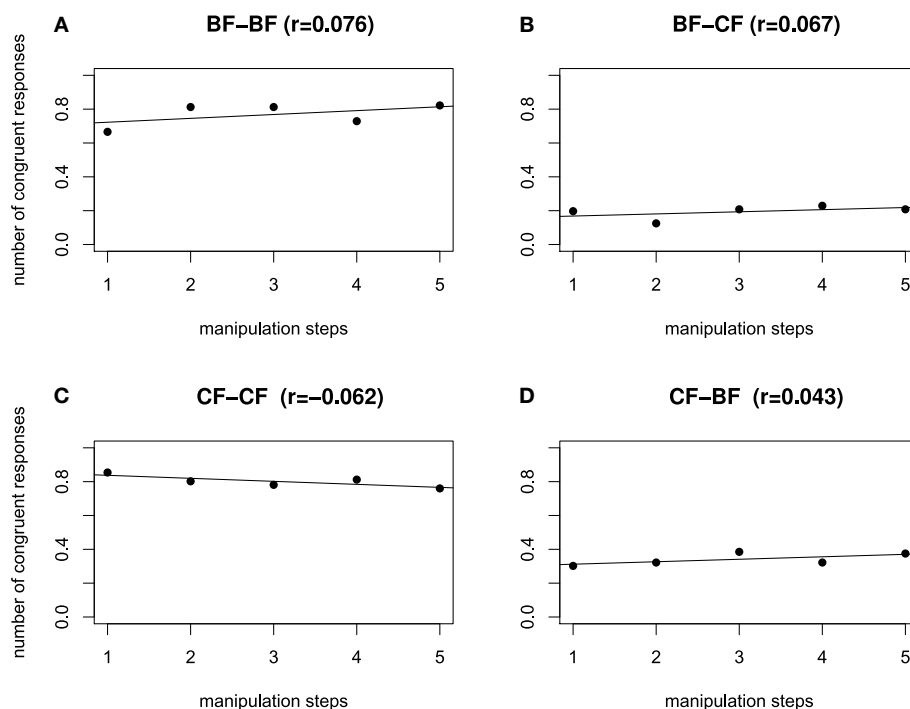
A series of semantic congruency experiments investigated the perceptual role of the phonetic differences found in



**TABLE 6 |** Report of the linear mixed effects model with the fixed factors context and manipulation and with congruent/incongruent ratings as dependent variable.

	Coefficients	SE	z-value	Sign.	Pr(>  z )
(Intercept)	0.0888	0.1736	0.512	n.s.	0.6089
context = CF	0.4521	0.2688	1.682	n.s.	0.0926
manipulation = step5	0.1771	0.1479	1.198	n.s.	0.2311
Interaction	−0.2227	0.1479	−1.506	n.s.	0.1321

n.s. refers to non significance.



**FIGURE 9 | Influence of L manipulation on the number of congruent responses, separated by dialog type, for all manipulation steps, starting from left to right with step 1: low L (150 Hz), step 2: (160 Hz), step 3: (170 Hz), step 4: (180 Hz), step 5: high L (190 Hz). Congruent dialog types BF-BF (A) and CF-CF (C), incongruent dialog types BF-CF (B) and CF-BF (D).**

the production experiment. The first perception experiment investigated whether listeners were able to perceive the phonetic differences found in production as a function of focus using congruent (BF-BF and CF-CF) and incongruent dialogs (BF-CF and CF-BF). Interestingly, the results of the perception study show that listeners are able to distinguish between congruent and incongruent dialogs, (see **Figure 3**) although the acoustic differences reported in **Table 2** were small. This might reveal that the overall shape of the intonation contour involves cues to perceive a contrastive or non-contrastive interpretation of an answer. As was shown in **Table 3**, prenuclear pitch accents in sentences containing a contrastive focus were realized lower on average before nuclear accents, while they were higher on average in case of broad focus sentences. This relation between the height of prenuclear and nuclear pitch accents seems to point to the fact that a nuclear rise-fall contour may be interpreted more global rather than locally at the nuclear pitch accent.

In order to investigate which parts of the rise-fall contour functionally interact with a contrastive interpretation, two separate perception experiments were conducted that examined whether the higher scaling of H\* accents causes the perceptual impression of contrastive focus, or whether the lower scaling of the low turning point is a sufficient phonetic cue. To this end, sentences with manipulated height values of the H\* peak, and of the low turning point were generated, respectively. The perception of the H\* accent manipulation revealed that a higher scaling of the H\* accent increased the perceptual impression of a contrastive accent. Specifically, contrastive contexts required

higher F0 values. Broad focus context allowed both, lower and higher H\* values, see (Féry and Kügler, 2008) for similar variations in speech production. Consequently, the free variation of upstepped accents (Féry and Kügler, 2008) and downstepped accents (Féry, 1993; Féry and Kügler, 2008; Grice et al., 2009) in broad focus contexts in speech production mirrors speech perception. The manipulation of the low F0 turning point, in turn, did not show an indication of a contrastive interpretation since the number of congruent responses did not change as a function of the low turning point value. The results appear to support the assumption that a contrastive focus compared to a broad sentence focus does not cause a different phonological category in German, but speak in favor of an interpretation that focus affects the pitch register (Féry and Kügler, 2008; Féry and Ishihara, 2010).

#### 4.1. The On-ramp vs. Off-ramp Debate

The experiments presented in this paper are partly related to the debate of how to analyse pitch accents, the so called “on-ramp” vs. “off-ramp” approach (Gussenhoven, 2004). The crucial assumption in the “off-ramp” approach is that the F0 movement **from** the pitch target is the essential of the pitch accent (off-ramp), whereas the “on-ramp” approach analyzes the F0 movement **toward** a pitch target as belonging to the pitch accent (on-ramp). The on-ramp approach is grounded in the ToBI tradition, which is “a system for transcribing the intonation patterns and other aspects of the prosody” of spoken utterances in a language variety (Beckman and Ayers-Elam, 1997). The

off-ramp approach was initiated by Gussenhoven (1984) and particularly studied in Hanssen et al. (2008) and Chen (2011).

Related to the present study, a rise-fall contour is phonologically analyzed as L+H\* L- in the “on-ramp” approach (Grice et al., 2005). The GToBI guidelines suggest to interpret a low turning point in F0 prior to the rise toward the accentual peak as a tone, while the perceptual impression of the stressed syllable is high (or rising). Hence, the rise is phonologically interpreted as a result of an F0 transition between a low leading tone (L+) and the accentual high tone (H\*) [cf. (2-a)].

From the off-ramp perspective, a rise-fall contour is analyzed as a phonological fall H\*+L following a phonetic rise (Féry, 1993; Grabe, 1998; Peters, 2014). The rise may vary in steepness and shape, but crucially it is not phonologically interpreted by means of a tone. With respect to the alignment of falling H\*+L pitch accents in German (Grabe, 1998) found that, in general, the position of the accentual peak is at the right edge of the accented syllable’s rime. Hence, there is an F0 transition toward the accentual peak, which however, is interpreted as a phonetic onglide that does not necessarily rise, or whose steepness may vary. Grabe (1998) carefully distinguished between non-final and final nuclear falling accents. Only in case of final falling accents, which are realized on a phrase-final accented syllable (e.g., [vɔlf] in *Ich bin der Wolf*. “I’m the wolf.” p. 73f), the peak position is realized earlier, that is at the onset of the accented vowel. This structural dependent variation of the accentual F0 peak led Grabe to conclude that the onglide only has phonetic properties since the onglide is less elaborated in the case of phrase-final accented syllables.

Similarly, structural conditions were found as evidence for an off-ramp analysis of Dutch prenuclear falling accents (Chen, 2011). In a comparison of prenuclear high and falling accents Chen (2011) observed a structural distinction rather than a functional one: independent of the information structural context (topic vs. focus) in which the accents were realized, the amount of sonorant segments within and after the accented syllable determined the accent pattern. If enough sonorant segments were present, a falling accent (H\*+L) was realized, if less sonorant material was present, a high rise (H\*) was realized. Similar to Grabe (1998), Chen (2011) concludes that the lack of a functional distinction of the two pitch accent types points to the fact that the distinction is phonetically motivated rather than phonologically determined.

As an alternative to the on-ramp and off-ramp interpretations of tonal contours, there are languages exhibiting tones that do not carry meaning, e.g., the accentual phrase tones in Tokyo Japanese (Gussenhoven, 2004), as opposed to a language like English where all post-lexical tones are supposed to carry meaning (Pierrehumbert and Hirschberg, 1990). On this note, the German rise-fall contour may constitute a case where the scaling of the accentual peak clearly contributes the interpretation of the contour with respect to contrastiveness while the rising part of the rise-fall contour does not contribute to this meaning, as our experiment three showed. Thus, a phonological interpretation of the rise-fall contour as L+H\* L- would be similar to the on-ramp approach except that contrary to the assumption proposed

in Grice et al. (2005), the leading low tone does not carry meaning.

Along these lines, our perceptual results of the manipulated stimuli may suggest that the onglide toward a high accentual F0 peak is either a phonetic transition (in the sense of the off-ramp approach) or a leading low tone that does not carry meaning. If the rise would have been a reflex of a phonological tone (L+) that carries a contrastive meaning as in English (Pierrehumbert and Hirschberg, 1990) native German listeners were expected to perceive this tone in the corresponding contexts. In particular, we were expecting a functional difference between a L+H\* accent and a simple H\* accent based on the assumption that L+H\* carries the meaning of contrast (Grice et al., 2005) given a similar functional distinction in English intonation (Pierrehumbert and Hirschberg, 1990; Beckman and Ayers-Elam, 1997). Manipulating the scaling of the onset of the rise (perception Experiment 3) did however not reveal that listeners relate a lower scaling to be congruent with a context that elicits contrast. We can thus conclude that a leading low tone does not seem to carry contrastive meaning in German.

## 4.2. Conclusion

This study investigated the phonetics of the rise-fall contour in German. In particular, it was tested whether phonetic differences in the rise-fall contour were realized in relation to contrastive and non-contrastive contexts, and which parts of the rise-fall contour seem to play a functional role in perception. The acoustic analysis of nuclear rise-fall contours elicited in broad and contrastive focus contexts revealed a significant difference for the realization of the accentual high tone, yet not for the low F0 turning point prior to the accentual high. In a series of semantic congruency perception tests, listeners judged the congruency of congruent and incongruent context-stimulus pairs on the basis of three different sets of stimuli: (i) original data from the production study in congruent contexts and cross-spliced yielding incongruent dialogs, (ii) stimuli with manipulated accentual high tone that were combined with originally congruent contexts and, again, cross-spliced with originally incongruent contexts, and (iii) stimuli with manipulated low F0 turning point of the rising part of rising-falling accent shapes, again combined with congruent and incongruent contexts. The first perception experiment revealed that listeners distinguish between nuclear rising-falling contours with respect to their focus context. The second perception experiment revealed that independent of stimulus origin, higher F0 peaks were rated significantly more frequent as congruent to contrastive focus contexts than lower peaks; hence, the scaling of the nuclear peak determined its contextual interpretation in our experiments as assumed in the literature on German intonation (Bannert, 1985; Alter et al., 2001; Braun, 2005, 2006; Baumann et al., 2006, 2007; Féry and Kügler, 2008; Grice et al., 2009; Sudhoff, 2010), and as argued by Gussenhoven (2004) in relation with the interpretation of focus in terms of the effort code. With respect to broad focus contexts, the results show that both upstepped and downstepped contours are rated as equally congruent reflecting a free variation of



the realization of the final (nuclear) accent in broad focus in speech production in German (Féry and Kügler, 2008). The third perception experiment revealed that manipulation of the low F0 turning point did not affect the perception as a function of focus context. Stimulus origin was rated more congruent than F0 manipulations.

The results of the perception experiments suggest that the scaling of the accentual peak is sufficient to license a contextual interpretation of a nuclear rising-falling accent shape (perception Experiment 2). The manipulation of a low F0 turning point prior to the accentual peak as a potential reflex of a low leading tone (L+) does not drive the perception as a function of focus context (perception Experiment 3). The results seem to support the view that focus affects the pitch register (Féry and Kügler, 2008; Féry and Ishihara, 2010), in our data a fact of pitch register raising of the nuclear accent peak. The production data also showed that the relation between prenuclear and nuclear accent peaks varies as a function of focus context. If the functional interpretation of pitch accents depends only on their local scaling, or if it is a matter of pitch accent relations within a sentence, or a combination thereof needs to be shown in future research.

## References

- Alter, K., Mleinek, I., Rohe, T., Steube, A., and Umbach, C. (2001). "Kontrastprosodie in Sprachproduktion und -perzeption," in *Kontrast: lexikalisch, Semantisch, Phonologisch*, Vol. 77, of *Linguistische Arbeitsberichte*, eds A. Steube and C. Umbach (Leipzig: Institut für Linguistik), 59–79.
- Bannert, R. (1985). Fokus, Kontrast und Phrasenintonation im Deutschen. *Z. Dialektol. Linguist.* 52, 289–305.
- Barker, G. S. (2002). *Intonation Patterns in Tyrolean German: An Autosegmental-Metrical Analysis*. Ph.D. thesis, University of California, Berkeley.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2013). *lme4: Linear Mixed-effects Models using Eigen and s4*, [R Package Version 1.0-4]. Available online at: <http://CRAN.R-project.org/package=lme4>
- Baumann, S. (2006). *The Intonation of Givenness. Evidence from German*. Tübingen: Niemeyer.
- Baumann, S., Becker, J., Grice, M., and Mücke, D. (2007). "Tonal and articulatory marking of focus in German," in *Proceedings of the 16th International Congress of Phonetic Sciences* (Saarbrücken), 1029–1032.
- Baumann, S., Grice, M., and Steindamm, S. (2006). "Prosodic marking of focus domains - categorical or gradient?," in eds R. Hoffmann and H. Mixdorff *Proceedings of Speech Prosody 2006* (Dresden: TUDpress), 301–304.
- Baumann, S., Mücke, D., and Becker, J. (2010). Expression of second occurrence focus in German. *Linguist. Ber.* 221, 61–78.
- Beckman, M. E., and Ayers-Elam, G. (1997). *Guidelines for ToBI Labelling*. Available online at: [http://www.ling.ohio-state.edu/research/phonetics/E\\_ToBI/singer\\_tobi.html](http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/singer_tobi.html)
- Bergmann, P. (2008). *Regionalspezifische Intonationsverläufe im Kölnischen: Formale und Funktionale Analysen Steigend-fallender Konturen*. Tübingen: Niemeyer.
- Boersma, P., and Weenink, D. (2013). *Praat: Doing Phonetics by Computer[Computer Program]*. Version 5.3.53. Available online at: <http://www.praat.org/>
- Braun, B. (2005). *Production and Perception of Thematic Contrast in German*. Frankfurt: Peter Lang Pub Inc.
- Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Lang. Speech* 49, 451–493. doi: 10.1177/00238309060490040201
- Büring, D. (2007). "Intonation, semantics and information structure," in *The Oxford Handbook of Linguistic Interfaces*, eds G. Ramchand and C. Reiss (Oxford: Oxford University Press), 445–474.
- Carlson, K. (2001). The effects of parallelism and prosody on the processing of gapping structures. *Lang. Speech* 44, 1–26. doi: 10.1177/00238309010440010101
- Chen, A. (2011). "What's in a rise: evidence for an off-ramp analysis of dutch intonation," in *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)* (Hong Kong), 448–451.
- Cummins, F., Doherty, C., and Dilley, L. (2006). "Phrase-final pitch discrimination in English," in *Proceedings of Speech Prosody* (Dresden: TUDpress), 467–470.
- del Mar Vanrell, M., Mascaró, I., Torres-Tamarit, F., and Prieto, P. (2013). Intonation as an encoder of speaker certainty: information and confirmation yes-no questions in Catalan. *Lang. Speech* 56, 163–190. doi: 10.1177/0023830912443942
- Dilley, L. (2005). *The Phonetics and Phonology of Tonal Systems*. Ph.D. thesis, Massachusetts Institute of Technology.
- Dilley, L., and Brown, M. (2007). Effects of pitch range variation on f0 extrema in an imitation task. *J. Phonet.* 35, 523–551. doi: 10.1016/j.wocn.2007.01.003
- Dilley, L., and Heffner, C. (2013). The role of F0 alignment in distinguishing intonation categories: evidence from American English. *J. Speech Sci.* 3, 3–67.
- Dilley, L. C. (2010). Pitch range variation in English tonal contrasts: continuous or categorical? *Phonetica* 67, 63–81. doi: 10.1159/000319379
- D'Imperio, M. (2000). *The Role of Perception in Defining Tonal Targets and Their Alignment*. Ph.D. thesis, The Ohio State University.
- Ellis, N. C. (2002). Frequency effects in language acquisition: a review with implications for theories of implicit and explicit language acquisition. *Stud. Second Lang. Acquis.* 24, 143–188. doi: 10.1017/S0272263102002024
- Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142, 27–41. doi: 10.1016/j.lingua.2012.12.006
- Féry, C. (1993). *German Intonational Patterns*. Tübingen: Niemeyer.
- Féry, C., and Ishihara, S. (2009). The phonology of second occurrence focus. *J. Linguist.* 45, 285–313. doi: 10.1017/S0022226709005702

## Acknowledgments

This research was supported by DFG-grants (KU 2323/1-2) to the project "Prosody in Parsing," principle investigators Frank Kügler, Caroline Féry and Shravan Vasishth, and projects D5 and T2 in the SFB 632 "Information structure" at Potsdam University. We are grateful to Tobias Guenther who provided considerable technical assistance, and to Jana Häussler for statistical advice. Parts of this work were presented at the 17. ICPhS, HongKong 2011 published as Kügler and Gollrad (2011), at the P&P-8 2012 in Jena, at the linguistics colloquium at Tübingen University in 2012, and as a poster at the DGfS Workshop "Prosody and Information Status in Typological Perspective" 2013 in Potsdam. We are grateful to the audiences for discussion. The paper has greatly benefited from the suggestions of and discussion with three reviewers.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01254>

- Féry, C., and Ishihara, S. (2010). "How focus and givenness shape prosody," in *Information Structure*, eds M. Zimmermann and C. Féry (Oxford: Oxford University Press), 36–63.
- Féry, C., and Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *J. Phonet.* 36, 680–703. doi: 10.1016/j.wocn.2008.05.001
- Forster, K. I., and Forster, J. C. (2003). Dmdx: a windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124. doi: 10.3758/BF03195503
- Fuchs, A. (1976). 'Normaler' und 'kontrastiver' Akzent. *Lingua* 38, 293–312.
- Gartenberg, R., and Panzlaff-Reuter, C. (1991). "Production and perception of  $F_0$  peak patterns in German," in *Studies in German Intonation*, Vol. 25 of *Arbeitsberichte (AIPUK)*, ed K. J. Kohler (Kiel: Institut für Phonetik und digitale Sprachverarbeitung), 29–113.
- Gilles, P. (2005). *Regionale Prosodie im Deutschen. Variabilität in der Intonation von Abschluss und Weiterweisung*. Berlin; New York: de Gruyter.
- Grabe, E. (1998). *Comparative Intonational Phonology. English and German*. Ph.D. thesis, Nijmegen University.
- Grice, M., and Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguist. Ber.* 191, 267–298.
- Grice, M., Baumann, S., and Benzmlüller, R. (2005). "German Intonation in Autosegmental-Metrical Phonology," in *Prosodic Typology, Chapter 2*, ed S.-A. Jun (Oxford: Oxford University Press), 430–458.
- Grice, M., Baumann, S., and Jagdfeld, N. (2009). Tonal association and derived nuclear accents—the case of downstepping contours in German. *Lingua* 119, 881–905. doi: 10.1016/j.lingua.2007.11.013
- Gussenhoven, C. (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.
- Gussenhoven, C. (1999). Discreteness and gradience in intonational contrasts. *Lang. Speech* 42, 283–305.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, C., and Rietveld, A. C. (1988). Fundamental frequency declination in Dutch: testing three hypotheses. *J. Phonet.* 16, 355–369.
- Gussenhoven, C., and Rietveld, T. (2000). The behavior of  $H^*$  and  $L^*$  under variations in pitch range in Dutch rising contours. *Lang. Speech* 43, 183–203. doi: 10.1177/00238309000430020301
- Hanssen, J., Peters, J., and Gussenhoven, C. (2008). "Prosodic effect of focus in Dutch declaratives," in *Proceedings of Speech Prosody 2008*, eds B. Bel and I. Marlien (Campinas), 609–612.
- Jackendoff, R. S. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.
- Klinghardt, H. (1925). *Sprechmelodie und Sprechakt*. Marburg: Elwert'sche Verlagsbuchhandlung.
- Klinghardt, H. (1927). *Übungen in Deutschem Tonfall*. Leipzig: Quelle und Meyer.
- Kohler, K. (1991). A model of German intonation. *Arbeitsber. Inst. Phonet. der Univ. Kiel* 25, 295–360.
- Kohler, K. J. (1987). "Categorical pitch perception," in *Proceedings of the 11th International Congress of Phonetic Sciences* (Tallin), 149–152.
- Krifka, M. (2008). Basic notions of information structure. *Acta Linguist. Hung.* 55, 243–276. doi: 10.1556/ALing.55.2008.3-4.2
- Kügler, F. (2007). *The intonational phonology of Swabian and Upper Saxon*. Tübingen: Niemeyer.
- Kügler, F. (2008). "The role of duration as a phonetic correlate of focus," in *Proceedings of the Speech Prosody 2008 Conference* (Campinas), 591–594.
- Kügler, F., Féry, C., and van de Vijver, R. (2003). "Pitch accent realization in German," in *Proceedings of the 15th International Congress of Phonetic Sciences*, eds M. J. Solé, D. Recasens, and J. Romero (Barcelona), 1261–1264.
- Kügler, F., and Gollrad, A. (2011). "Production and perception of contrast in German," in *Proceedings of the 17th International Congress of Phonetic Sciences* (Hong Kong: ICPHS XVII), 1154–1157.
- Ladd, D., and Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *J. Phonet.* 25, 313–342.
- Ladd, D. R. (2008). *Intonational Phonology, 2nd Edn*. Cambridge: CUP.
- Ladd, D. R., and Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: experimental evidence. *J. Phonet.* 31, 81–112. doi: 10.1016/S0095-4470(02)00073-6
- Ladd, D. R., Verhoeven, J., and Jacobs, K. (1994). Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects. *J. Phonet.* 22, 87–99.
- Mayer, J. (1997). *Intonation und Bedeutung*, Ph.D. thesis, Universität Stuttgart, [Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung, AIMS 3(4)].
- Molnár, V. (2002). Contrast from a contrastive perspective. *Lang. Comput.* 39, 147–161.
- Nash, R., and Mulac, A. (1980). "The intonation of verifiability," in *The Melody of Language, Chapter 17*, eds L. R. Waugh and C. H. van Schooneveld (Baltimore: University Park Press), 219–241.
- Niebuhr, O. (2007). *Perzeption und Kognitive Verarbeitung der Sprechmelodie: Theoretische Grundlagen und Empirische Untersuchungen*. Berlin: Walter de Gruyter.
- Niebuhr, O., and Kohler, K. J. (2004). "Perception and cognitive processing of tonal alignment in German," in *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, ed M. Bel (Beijing), 155–158.
- Peters, J. (2005). *Intonation*, in *Duden-Die Grammatik, Chapter 2*. Mannheim: Bibliographische Institut Mannheim, Dudenverlag.
- Peters, J. (2006). *Intonation Deutscher Regionalsprachen*. Berlin; New York: de Gruyter.
- Peters, J. (2014). *Intonation*. Heidelberg: Winter.
- Petrone, C., and Niebuhr, O. (2014). On the intonation of German intonation questions: the role of the prenuclear region. *Lang. Speech* 57, 108–146. doi: 10.1177/0023830913495651
- Phéby, J. (1975). *Intonation und Gramatik im Deutschen*. Berlin: Akademie Verlag.
- Pierrehumbert, J. (2001). "Exemplar dynamics: word frequency, lenition, and contrast," in *Frequency effects and the emergence of lexical structure*, eds J. Bybee and P. Hopper (Amsterdam: John Benjamins), 137–157.
- Pierrehumbert, J., and Hirschberg, J. (1990). "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication*, eds P. R. Cohen, J. Morgan, and M. E. Pollack (Cambridge, MA: MIT Press), 271–311.
- Pierrehumbert, J., and Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica* 46, 181–196.
- Post, B. (2000). *Tonal and Phrasal Structures in French Intonation*. Ph.D. thesis, Nijmegen University.
- Prieto, P. (2012). "Experimental methods and paradigms for prosodic analysis," in *The Oxford Handbook of Laboratory Phonology*, eds A. Cohn, C. Fougeron, and M. Huffman (Oxford: Oxford University Press), 528–537.
- Rathcke, T., and Harrington, J. (2010). "The variability of early accent peaks in Standard German," in *Laboratory Phonology*, Vol. 10, eds C. Fougeron, B. Kühnert, M. d'Imperio, and N. Vallée (Berlin: De Gruyter Mouton), 533–555.
- Redi, L. (2003). "Categorical effects in production of pitch contours in English," in *Proceedings of the 15th International Congress of Phonetic Sciences*, eds M. J. Solé, D. Recasens, and J. Romero (Barcelona), 2921–2924.
- Remijsen, B., and van Heuven, V. J. (1999). "Gradient and categorical pitch dimensions in Dutch: diagnostic test," in *Proceedings of the 14th International Congress of Phonetic Sciences* (San Francisco, CA), 1865–1868.
- Repp, S. (2010). Defining 'contrast' as an information-structural notion in grammar. *Lingua* 120, 1333–1345. doi: 10.1016/j.lingua.2009.04.006
- Rietveld, A., and Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *J. Phonet.* 13, 299–308.
- Schneider, K., and Lintfert, B. (2003). "Categorical Perception of Boundary Tones in German," in *Proceedings of the 15th International Congress of Phonetic Sciences*, eds M. J. Solé, D. Recasens, and J. Romero (Barcelona), 631–634.
- Schweitzer, K., Walsh, M., Calhoun, S., Schütze, H., Möbius, B., Schweitzer, A., et al. (2015). Exploring the relationship between intonation and the lexicon: evidence for lexicalised storage of intonation. *Speech Commun.* 66, 65–81. doi: 10.1016/j.specom.2014.09.006
- Selkirk, E. (1984). *Phonology and Syntax. The Relation between Sound and Structure*. Cambridge: MIT Press.
- Sudhoff, S. (2010). Focus particles and contrast in German. *Lingua* 120, 1458–1475. doi: 10.1016/j.lingua.2009.02.011
- Toepel, U., Pannekamp, A., Alter, K., Hahne, A., and Friederici, A. D. (2005). Prosody-driven sentence processing: an event-related brain potential study. *J. Cogn. Neurosci.* 17, 407–421. doi: 10.1162/0898929053279450

- Truckenbrodt, H. (2005). A short report on intonation phrase boundaries in German. *Linguist. Ber.* 203, 273–296.
- Truckenbrodt, H. (2007). “Upstep on edge tones and on nuclear accents,” in *Tones and Tunes*. Vol. 2: *Experimental Studies in Word and Sentence Prosody*, eds C. Gussenhoven and T. Riad (Berlin: Mouton), 349–386.
- Truckenbrodt, H. (2012). “Semantics of intonation,” in *Semantics. An International Handbook of Natural Language Meaning*, eds C. Maienborn, K. V. Heusinger, and P. Portner (Berlin: De Gruyter), 2039–2969.
- Uhmann, S. (1991). *Fokusphonologie: Eine Analyse Deutscher Intonationskonturen im Rahmen der Nicht-linearen Phonologie*. Tübingen: Niemeyer.
- von Essen, O. (1964). *Grundzüge der deutschen Satzintonation*, 1st Edn. 1956. Ratingen: Henn.
- Wunderlich, D. (1988). “Der Ton macht die Melodie - Zur Phonologie der Intonation des Deutschen,” in *Intonationsforschungen*, ed H. Altmann (Tübingen: Niemeyer), 1–40.
- Yang, Z., Ramanarayanan, V., Byrd, D., and Narayanan, S. S. (2013). “The effect of word frequency and lexical class on articulatory-acoustic coupling,” in *Proceedings of Interspeech* (Lyon), 973–977.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Kügler and Gollrad. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## ADVANTAGES OF PUBLISHING IN FRONTIERS



### FAST PUBLICATION

Average 90 days  
from submission  
to publication



### COLLABORATIVE PEER-REVIEW

Designed to be rigorous –  
yet also collaborative, fair and  
constructive



### RESEARCH NETWORK

Our network  
increases readership  
for your article



### OPEN ACCESS

Articles are free to read,  
for greatest visibility



### TRANSPARENT

Editors and reviewers  
acknowledged by name  
on published articles



### GLOBAL SPREAD

Six million monthly  
page views worldwide



### COPYRIGHT TO AUTHORS

No limit to  
article distribution  
and re-use



### IMPACT METRICS

Advanced metrics  
track your  
article's impact



### SUPPORT

By our Swiss-based  
editorial team