frontiers in
PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 23 March 2012
doi: 10.3389/fpsyg.2012.00070

Are the products of statistical learning abstract or stimulus-specific?

Athena Vouloumanos[1]*, Patricia E. Brosseau-Liard[2], Evan Balaban[3] and Alanna D. Hager[4]

[1] Department of Psychology, New York University, New York, NY, USA
[2] Department of Psychology, University of British Columbia, Vancouver, BC, Canada
[3] Department of Psychology, McGill University, Montreal, QC, Canada
[4] Department of Psychology, University of Victoria, Victoria, BC, Canada

Edited by:
Marcela Pena, Catholic University of Chile, Chile

Reviewed by:
LouAnn Gerken, University of Arizona, USA
Padraic Monaghan, Lancaster University, UK
Luca Lorenzo Bonatti, Universitat Pompeu Fabra, Spain

***Correspondence:**
Athena Vouloumanos, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA.
e-mail: athena.vouloumanos@nyu.edu

Learners can segment potential lexical units from syllable streams when statistically variable transitional probabilities between adjacent syllables are the only cues to word boundaries. Here we examine the nature of the representations that result from statistical learning by assessing learners' ability to generalize across acoustically different stimuli. In three experiments, we compare two possibilities: that the products of statistical segmentation processes are abstract and generalizable representations, or, alternatively, that products of statistical learning are stimulus-bound and restricted to perceptually similar instances. In Experiment 1, learners segmented units from statistically predictable streams, and recognized these units when they were acoustically transformed by temporal reversals. In Experiment 2, learners were able to segment units from temporally reversed syllable streams, but were only able to generalize in conditions of mild acoustic transformation. In Experiment 3, learners were able to recognize statistically segmented units after a voice change but were unable to do so when the novel voice was mildly distorted. Together these results suggest that representations that result from statistical learning can be abstracted to some degree, but not in all listening conditions.

Keywords: speech perception, representation, generalization, segmentation, acoustics, statistical learning

INTRODUCTION

One of the first tasks for a novice language-learner is to extract words from what is essentially a continuous stream of sound. Whereas written text provides spaces between words to mark word boundaries, there are no clear acoustic cues in speech input as to the location of word boundaries (e.g., Cole et al., 1980). Learners must thus generate hypotheses about where possible boundaries between words might be. At the same time, the problem is made especially challenging because naturally occurring speech signals in the environment are acoustically variable. As Church and Fisher (1998, p. 523) put it, "Locating and identifying words in continuous speech presents a number of well-known perceptual problems. These problems fall into two interrelated classes: word segmentation and acoustic/phonetic variability." The current paper explores the intersection of these two problems of acquisition by asking whether a segmentation process that tracks statistically variable transitional probabilities across adjacent elements could help a learner segment potential words across acoustically variable environments.

Full-fledged words are rich, abstract, lexical entries that speakers of a language can recognize across contexts, across speakers, and across pronunciations. Representing speech sound patterns as potential lexical units is important for identifying new occurrences of these words, and essential for eventually mapping these words onto meanings. Here, we examine the nature of the representations extracted by statistical segmentation processes. We consider whether statistical segmentation processes produce "presemantic" sound representations that behave like lexical units or "acoustic" representations that are stimulus-bound. Since statistical segmentation studies to date have largely used acoustically identical sounds during training and testing, the nature of the lexical units learned from statistical segmentation processes remains to be examined.

Though learners use various sources of linguistic information – for example, prosodic (Johnson and Jusczyk, 2001), phonotactic (Mattys and Jusczyk, 2001), and allophonic (Jusczyk et al., 1999) – to segment words from the speech stream, a statistical process that keeps track of statistical regularities (e.g., transitional probabilities) between sounds might play a role in word segmentation (Saffran et al., 1996a). Using such a statistical mechanism, adjacent sounds that occur sequentially with a higher transitional probability would be segmented as part of the same word, while transitions with low or near-zero probability would likely mark boundaries between different words. Tracking statistically variable transitional probabilities has the advantage of being universally available for acquisition, unlike phonotactic constraints or prosodic patterns which require previous language-specific knowledge.

Both adults (Saffran et al., 1996b) and infants (Saffran et al., 1996a) can use transitional probabilities to segment a speech stream. However, statistical learning mechanisms appear to be neither domain-specific, nor specific to humans. In addition to speech sounds, this general-purpose learning mechanism can track probabilities in tone sequences and musical timbres (Saffran et al., 1999; Tillmann and McAdams, 2004) and visual patterns (Fiser and Aslin, 2002; Kirkham et al., 2002; Turk-Browne et al., 2005) and thus is not specific to language learning. It is also not exclusive to humans, as cotton-top tamarin monkeys and rodents can also learn from statistical regularities in speech input (Hauser et al., 2001; Newport et al., 2004; Toro and Trobalon, 2005). Its availability for segmenting non-speech and its use by non-humans brings into question the lexical nature of the products of statistical learning.

The starting point to our investigation is the fact that speech is an extremely variable acoustic signal. For instance, changes in speaking rate affect different phonemes differently (Jusczyk and Luce, 2002). Similarly, coarticulation between adjacent speech sounds renders every instance of a word acoustically different depending on the sounds that precede and follow it (Curtin et al., 2001). Variability can be detrimental to speech processing. In adults, variability has a detrimental effect on speech perception, especially under distorted conditions or in low-probability phrases (Mullennix et al., 1989). More variability in a familiarization stream of speech leads to more encoding of acoustic details and a larger influence of acoustic properties on later performance (Ju and Luce, 2006). When acoustic variation is introduced, infants perform less well than when the acoustic content is constant (Grieser and Kuhl, 1989; Swingley and Aslin, 2000). Other studies suggest that 2-month-old infants can perceptually normalize across talkers but also encode the difference between the voices of different talkers, and that they take longer to habituate to a phoneme uttered by several talkers than to the same phoneme uttered by a single talker (e.g., Jusczyk et al., 1992). Variability therefore seems generally detrimental to speech perception except in some cases where variability can improve performance (e.g., Goldinger et al., 1991).

Though variability in the speech signal generally results in less efficient and less accurate speech processing (Mullennix et al., 1989; Van Tasell et al., 1992), humans are able to ignore or compensate for a great degree of natural variation and artificial distortion in their use of spoken language. Adults and infants readily ignore incorrect or incomplete information in the acoustic realization of words. For example, replacing a phoneme with a noise burst does not impair adults' recognition of words, and often the noise burst goes undetected (Warren, 1970). Even infants are able to recognize a familiar word in which the initial phoneme has been changed, which suggests that their representations of familiar words are not tied to specific sounds within the words (Swingley and Aslin, 2002). A particularly strong example of how little impact acoustic variation can have on speech recognition is that of whispered speech. Samuel (1988) showed that selective adaptation occurs for phonemes across normal and whispered speech, despite large acoustic differences in the phonemes' acoustic properties between the two contexts. Words, or lexical units, are thus partly represented abstractly allowing speakers to recognize the same lexical unit across natural variability in different speakers, contexts, and pronunciations.

Human speech perception is robust even when the temporal or spectral properties of speech have been distorted (for reviews, see Scott and Johnsrude, 2003; Davis and Johnsrude, 2007). Saberi and Perrott (1999) found that intelligibility remains almost perfect even when words have been severely distorted: comprehensible English sentences were divided into segments that were then either locally time-reversed, or temporally delayed. Time-reversals of segments up to 50 ms in length were identified perfectly accurately, with accuracy decreasing with longer segments, reaching chance performance at around 130 ms reversals. Time delays were much less detrimental to performance. A spectral manipulation called band pass-filtering, which removes much spectral information from speech including formant transitions (Stickney and Assmann, 2001) can also result in highly intelligible speech with as few as three or four bands when slowly varying temporal information is available (Shannon et al., 1995; Davis et al., 2005; Hervais-Adelman et al., 2008). Despite extreme variability in the signal, adults are able to understand speech in a consistent way and recognize the same word in different contexts. Although lexical representations are not fully abstract, instead representing acoustic information specific to individuals and to contexts (Goldinger, 1996, 1998; Kraljic and Samuel, 2005; Werker and Curtin, 2005), adults' proficiency at recognizing the same lexical unit in different acoustic contexts requires that humans draw on an abstract linguistic representation that is in part independent of specific acoustic properties.

Here, we use variability as a tool to understand the products of statistical learning. Although humans can use statistical regularities in fluent speech to readily acquire information about possible lexical units, little is known about how learners represent these newly segmented units, specifically whether segmented units are coded as acoustic units or as putative lexical entries. Only a few previous studies bear directly on this issue. Studies in which infants were trained on nonsense streams presented either in infant-directed speech (Thiessen et al., 2005) or with stressed syllables (Thiessen and Saffran, 2003) showed that infants were able to recognize segmented "words" from the "part-words" when these were subsequently presented in adult-directed or unstressed speech. Although this suggests some degree of abstraction in the representation of segmented speech units, the higher pitch and greater length of infant-directed speech and stressed syllables typically attract infant attention (e.g., Fernald, 1985), and this increased attention could account for infants' better performance (see also footnote 1 in Thiessen and Saffran, 2003 reporting higher attrition to the monotone training). Similarly, adults can recognize shorter test tokens after segmentation training on lengthened tokens (Saffran et al., 1996b). Syllable lengthening, however, is present in natural speech (especially in word-final or sentence-final positions; Klatt, 1975) and thus adults in these studies may have normalized acoustic differences with which they were already familiar. More recently, infants and adults were shown to be able to map newly segmented words onto novel objects in a word learning task (Graf Estes et al., 2007; Mirman et al., 2008). Although both adults and infants more easily mapped the high probability syllable–strings onto referents, consistent with their representation as lexical units better performance may have also been due to better encoding and recognition of the high probability syllable–string simply due to its higher probability. Much like the process of lexicalization of new words (Gaskell and Dumay, 2003), initial encoding of novel syllable–strings as putative words may have relied on a stimulus-linked memory trace that is not actually integrated into the lexicon. These studies do not directly choose between abstract and acoustic representations.

In the domain of artificial grammar learning, which requires computations across statistically regular units, results conflict: some studies point to abstract representations (Altmann et al., 1995), and others stimulus-specific representations (Conway and Christiansen, 2006). Perhaps the best evidence for representing statistically regular information in an abstract and generalizable format comes from the visual domain. Turk-Browne et al. (2005) demonstrated that adult learners presented with statistical regularities instantiated in colored visual stimuli (novel green or red shapes presented sequentially), were able to abstract these regularities to black shapes during test. This suggests that products of statistical learning, at least in the visual domain, can be generalized to perceptually different stimuli. At the same time, however, studies which examined whether statistical learning transfers across modalities found that the products of statistical learning were stimulus-specific (Conway and Christiansen, 2006). Whether statistical learning in the auditory domain can be generalized to perceptually different stimuli remains to be tested. In this paper, we take a complementary approach to investigating how the output of statistical learning processes is represented by focusing on adults' ability to restore acoustically variable and distorted speech.

In three experiments, we investigate how auditory units segmented through statistical processes are represented and abstracted by training and testing adults on acoustically dissimilar units. The acoustic characteristics of the sounds were temporally distorted by locally time-reversing the speech (see **Figure 1**). Previous studies show that reversals of 50 ms result in near perfect intelligibility, whereas reversals of 100 ms only allow listeners to recover 50% of the utterance. Performance is linearly related to the duration of the reversed segment (Saberi and Perrott, 1999). Time-reversal distorts the temporal envelope and the fine structure of the spectrum. Much acoustic evidence for place, manner, and voicing of speech segments is conveyed in temporal regions where the speech spectrum is changing rapidly, that is, when the vocal tract opens and closes rapidly to produce consonants (Stevens, 1980; Liu, 1996). At the same time, because consonants are briefer in duration and lower in intensity than vowels, which have a longer spectral steady state, they are less robust than vowels and more vulnerable to distortion (e.g., Miller and Nicely, 1955; Assmann and Summerfield, 2004). Time-reversals therefore affect the speech intelligibility of consonants more than vowels.

Adults were familiarized with a series of trisyllabic nonsense words that were concatenated into 10-min streams of continuous speech, in which transitional probabilities between adjacent segments were the only cues to word boundaries (e.g., Saffran et al., 1996a). Familiarization consisted of either normal speech, or the same speech stream, temporally distorted by locally time-reversing 50-ms or 100-ms segments (Saberi and Perrott, 1999), giving the sounds a reverberant quality. Participants were subsequently tested on their ability to discriminate the familiar "words" from trisyllabic "non-words" whose component syllables were drawn from the familiarization stream but which had been subjected to acoustic transformations (temporal reversals, or a voice change to a different gender). We used non-words rather than part-words because evidence for learning was more robust with non-words (e.g., Saffran et al., 1996b) providing a starting point for examining the influence of acoustic distortions. Our goal was to determine whether products of segmentation through statistical learning are tied to the specific acoustic properties of the training environment or if the units are represented more abstractly, allowing newly acquired words to be recognized when they differ acoustically.

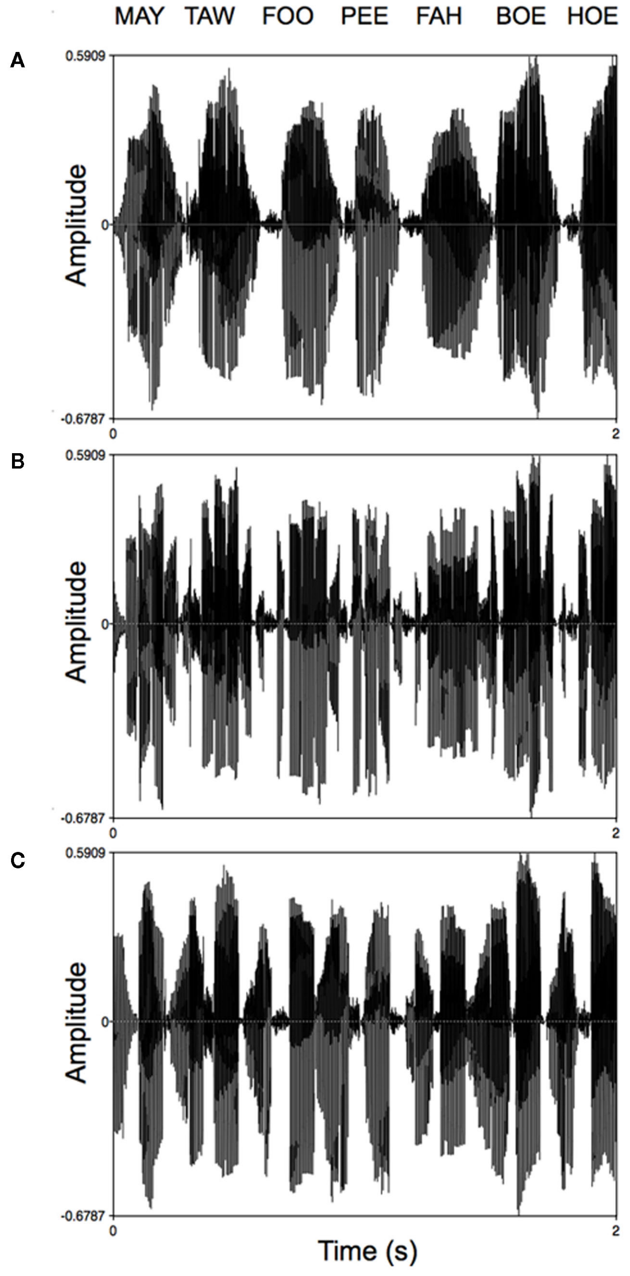

FIGURE 1 | A 2-s excerpt from (A) normal, (B) 50-ms time-reversed, and (C) 100-ms time-reversed familiarization streams.

EXPERIMENT 1: REPRESENTING STATISTICAL PRODUCTS: ABSTRACTING TO DISTORTED SPEECH

In Experiment 1A, we replicate the findings of previous statistical learning studies (e.g., Saffran et al., 1997) using different trisyllabic words and a shorter familiarization stream (similar to Peña et al., 2002). In Experiment 1B, participants were familiarized with the same speech stream as 1A, and then tested on words that were time-reversed every 50 ms, a type of distortion that is not severe enough to impair recognition of known English words (Saberi and Perrott, 1999). In Experiment 1C, participants were tested with words time-reversed every 100 ms, which is a more drastic form of distortion that has been shown to impair speech recognition (Saberi and Perrott, 1999).

MATERIALS AND METHODS

Participants

Forty-five participants were recruited through the undergraduate participant pool, or through ads posted on the university campus. Participation in the study was compensated with either course credit in one psychology course or $5. Participants gave informed consent and reported being fluent in English and having no hearing impairment. Participants were assigned to either condition A (15), condition B (15), or condition C (15). All procedures were approved by the research ethics board at McGill University.

Stimuli

The experimental materials consisted of 12 consonants (b, d, f, g, h, k, l, m, p, r, t, w) and 6 vowels (/ah/, /ay/, /aw/, /oh/, /oo/, /ee/) combined to form 18 consonant–vowel syllables (see **Table 1**). Syllables were generated individually in DECtalk 5.0 (Fornix Corporation, UT, USA). The settings for the electronic female voice "Betty" were used, with a "speech rate" (not related to syllable rate) setting of 205 words per minute, with a richness of 0, a smoothness of 70%, average pitch of 180 Hz and pitch range of 0. Syllables were 228–388 ms (average = 296 ms) in length (see **Table 1** for details.) These syllables were combined into six trisyllabic units, hereafter called "words," which were concatenated into four different 10-min speech streams. Only two streams were used in conditions A, and B, whereas all four streams were used in condition C. The order of words in each of the streams was randomized, with the only constraint being that the same word could not occur twice in a row. Transitional probabilities between syllables within a word were 1.0, but only 0.2 for syllables spanning word boundaries (since each word in the stream is followed randomly by one of the other five words). For 1A, and 1B the amplitude of the speech stream was attenuated every 50 ms with a 2.5-ms decrease in amplitude and a 2.5-ms increase in amplitude. For 1C, amplitude was attenuated every 100-ms with a 2.5-ms decrease in amplitude and a 2.5-ms increase in amplitude. Attenuations at reversal boundaries prevented noise bursts that come from abrupt changes in amplitude in artificially reversed speech and change the properties of the percept.

The test trials were composed of the six trisyllabic words presented in the familiarization streams, as well as six trisyllabic non-words which consisted of the same syllables as the words, but in new orderings, such that transitional probabilities between each syllable of these non-words was 0, compared to 1.0 transitional probabilities for syllables within the words. We opted to use

Table 1 | The word and non-words used in the three experiments and thee lengths of the component syllables in ms.

Syllable 1	ms	Syllable 2	ms	Syllable 3	ms	Word
WORDS						
Pee	229	Fah	388	Boe	267	Peefahboe
Roo	304	Wee	272	Law	336	Rooweelaw
Hoe	329	Lay	298	Gee	228	Hoelaygee
Kaw	298	Goo	267	Pah	306	Kawgoopah
May	286	Taw	298	Foo	330	Maytawfoo
Dah	300	Koe	285	Way	298	Dahkoeway
	291.0		301.3		294.2	Average length
Syllable 1	**ms**	**Syllable 2**	**ms**	**Syllable 3**	**ms**	**Non-word**
NON-WORDS						
Lay	298	Hoe	329	Pee	229	Layhoepee
Wee	272	Kaw	298	Roo	304	Weekawroo
Boe	267	Gee	228	Taw	298	Boegeetaw
Goo	267	Dah	300	May	286	Goodahmay
Law	336	Way	298	Fah	388	Lawwayfah
Pah	306	Foo	330	Koe	285	Pahfookoe
	291.0		297.2		298.3	Average length

The six non-words were created with the following constraints: half began with a medial syllable and half began with a final syllable. No syllable occurred in the same position in the non-words as it did in the words (e.g., if "Wee" was in medial position in a word, it would not occur medially in a non-word) so as to prevent learners from falsely recognizing a nonsense word from positional encoding of its component syllables. Finally, each of the vowels occurred once in each position in the words and once in each position in the non-words but was never repeated within a given word, or given non-word.

non-words (as in Saffran et al., 1996a, Experiment 1; Saffran et al., 1996b, Experiment 1; Saffran et al., 1997, Experiments 1 and 2), rather than part-words, because we were interested in whether learners could abstract to new materials and not the specifics of which transitional probabilities – trigrams or bigrams – learners were using. Learners may have extracted either bigrams or trigrams from the trainings streams. In 1A, words and non-words were acoustically identical to the training phase, replicating previous statistical learning studies (Saffran et al., 1997). In 1B, the same six words and six non-words were time-reversed every 50-ms. In 1C, the six words and six non-words were time-reversed every 100-ms[1].

Apparatus and procedure

Participants were tested individually in a small quiet room, using an Apple G4 400 MHz computer. All participants were familiarized with one of the 10-min speech streams played over two speakers placed on either side of the monitor using QuickTime (Apple, Cupertino, CA, USA). They were instructed to actively listen to the stream, but not to try to memorize the stream or do anything other than simply listening.

To create the test pairs, each of the six words was paired with each of the six non-words exhaustively for 36 total test trials. Each particular word was played first in three test trials and second in three test trials. In the test phase, participants were randomly presented with 36 pairs of words and non-words, and asked to choose which of the pair seemed most familiar. In half of the 36 trials, words were presented first, and in the other half, non-words were presented first. For participants in 1A, these words and non-words were normal, participants in 1B were tested on 50-ms time-reversed materials, and participants in 1C were tested on the 100-ms time-reversed materials. Guessing was encouraged if participants were not sure of the answer. Stimuli were presented and responses recorded using PsyScope 1.2.5 (Cohen et al., 1993). After each pair was presented a short pause allowed participants to press the "z" key on the keyboard, labeled "1," if they believed that first trisyllabic sequence to be most familiar, and to press the "m" key, labeled "2," if they believed the second sequence to be most familiar. The number of correct responses in the normal and reversed test conditions was then calculated. For three participants, only 35 trials were recorded, and thus the percentage of correct responses was scored out of 35. The entire session lasted approximately 20 min.

RESULTS AND DISCUSSION

Preliminary analyses were conducted in order to assess whether there was any difference in performance caused by particular familiarization streams. Since no significant effect was found, the different streams were combined in subsequent analyses. In 1A, words were chosen over non-words on average in 64.5% of test trials, reliably above chance [all t-tests are two-tailed one-sample tests; $t(14) = 7.05$, $p < .001$; see **Figure 2**][2]. In 1A, we thus replicated the findings of previous statistical learning studies by demonstrating that adults can segment words using statistically occurring transitional information (Saffran et al., 1997).

In Experiments 1B and 1C, we examined whether the type of representation segmented from a statistical speech stream allows learners to recognize words when they have been acoustically distorted by 50 and 100-ms time-reversals. In Experiment 1B, with 50-ms time-reversed materials, words were chosen over non-words in 61.3% of test trials. A planned two-tailed one-sample t-test indicated that this proportion was significantly different from chance [$t(14) = 5.61$, $p < .001$]. In Experiment 1C, 100-ms time-reversed words were selected on average in 54.4% of trials, significantly above chance [planned two-tailed one-sample t-test $t(14) = 3.84$, $p = .002$]. Performance differed between conditions, $F(2,42) = 8.39$, $p = .001$, with performance on 100-ms reversals being reliably worse than normal words (LSD Mean Difference = 10.1, $p < .001$) and 50-ms reversals (LSD Mean Difference = 6.96, $p = .009$). Learners were able to recognize segmented words under both distortion conditions, although performance was worse with larger distortion.

Adults were able to restore speech sounds that were acoustically distorted and to map them to their representation of newly segmented units; however, restoration became more difficult as the amount of acoustic distortion increased. Interestingly, adults' performance in restoring newly learned segmented novel units parallels their ability to restore *known* speech which includes higher-level syntactic and semantic cues (Saberi and Perrott, 1999). This suggests that syntactic and semantic information is not essential for speech restoration, since our speech units were nonsense words presented in isolation, but that statistical regularities can suffice for generalization. The products of statistical learning are abstract enough to allow learners to recognize newly segmented units when they have been acoustically distorted.

EXPERIMENT 2: ACQUIRING PUTATIVE LEXICAL UNITS FROM DISTORTED STATISTICAL SPEECH STREAMS

Adults have surprisingly robust abilities to recover *known* words from speech signals distorted both in frequency (Blesser, 1972; Remez et al., 1981; Shannon et al., 1995) and in time (e.g., Saberi and Perrott, 1999; Greenberg and Arai, 2004). In Experiment 1 we showed that listeners can recognize newly segmented speech units which have been altered with an artificial temporal distortion, suggesting that listeners can perceptually restore newly

[1] In order to ascertain that selection of words over non-words in the test phase would be due to statistical learning from the familiarization stream and not to characteristics intrinsic to our stimuli, 18 additional participants were presented with the test phase without any familiarization phase. Each of these participants was tested on pairs of words and non-words that were normal (as in **Figure 1A**), 50-ms time-reversed (as in **Figure 1B**), and 100-ms time-reversed (as in **Figure 1C**). The order of these three test phases was counterbalanced across participants. As predicted, performance was at chance for all levels of distortion [for normal words, mean percent selection of words = 49.9%, $t(17) = 0.106$, p = ns; for 50-ms reversed: mean = 50.9%, $t(17) = 0.377$, p = ns; for 100-ms reversed: mean = 48.3%, $t(17) = 0.821$, p = ns]. This indicates that results from our experiments are the product of learning that occurs in the familiarization phase and not intrinsic stimulus characteristics.

[2] Sixteen additional participants were familiarized on streams which were attenuated every 100-ms. After this familiarization, words were selected as more familiar than non-words on average 64.6% of the time, indistinguishable from performance in Experiment 1A, confirming that the position of the attenuation within the syllable stream (the 2.5-ms decrease in amplitude and 2.5-ms increase in amplitude between every time-reversal) had no effect on performance.

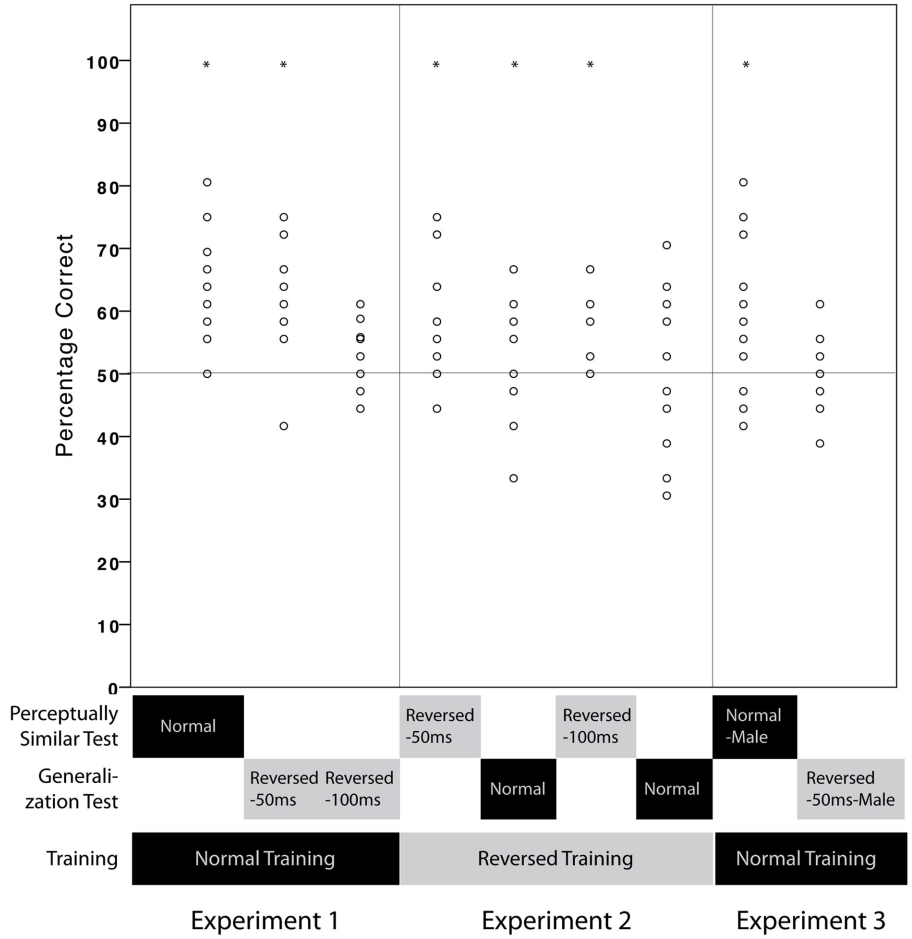

FIGURE 2 | Mean percentage correct for participants in each Experimental condition. Training conditions are indicated in the lower boxes: normal training and test conditions are indicated in light text on a dark frame. Reversed training and test conditions are indicated in dark text on a light frame. Test conditions are indicated in the upper boxes: Perceptually Similar Test conditions are ones in which the training and test stimuli are perceptually similar. Generalization Test conditions are ones in which the training and test stimuli are perceptually distinct.

segmented speech material. However, other language learning processes show some dissociation between mechanisms involved in acquiring knowledge (e.g., acquiring a rule is easier from speech than non-speech; Marcus et al., 2007) and mechanisms involved in using knowledge (e.g., rules are applied equally to speech and non-speech; Marcus et al., 2007). In Experiment 2 we examine the extent to which humans can *acquire* and generalize novel lexical forms under distorted listening conditions.

Previous studies show that adult listeners can selectively adapt to certain types of acoustic distortions in a process called perceptual learning, however, perceptual learning is greatly facilitated when listeners are given feedback on sentences containing known words (e.g., Davis et al., 2005). Since the current study contains neither syntactic nor semantic cues, it is not clear whether language learners will be able to segment units under distorted conditions.

This experiment investigates whether statistical learning can occur with distorted input, and, if so, whether the representation thereby acquired will be equally resistant to acoustic change. In Experiment 2A, participants were familiarized with a 10-min 50-ms time-reversed speech stream; half of them were tested on 50-ms time-reversed words which maintain the acoustic distortion, and the other half were tested on non-distorted words that mirror the amount of acoustic transformation in Experiment 1B. In Experiment 2B, participants were familiarized with a 10-min 100-ms time-reversed stream; half of them were tested on 100-ms time-reversed words, maintaining the amount of distortion, and the other half were tested on non-distorted words, thus mirroring the amount of acoustic transformation from Experiment 1C.

MATERIALS AND METHODS

Participants

Sixty-two participants were recruited through the undergraduate participant pool and through ads posted on campus. Participants gave informed consent, reported being fluent in English and having no hearing impairment. They were compensated with either course credit in one psychology course or $5. In 2A, after 50-ms time-reversed familiarization, participants were assigned to one of the two test conditions, 50-ms reversed test

(15) or non-distorted test (15). In 2B, after 100-ms time-reversed familiarization, participants were assigned to either the 100-ms reversed test (16) or the non-distorted test (16). Four additional participants were tested and excluded from the sample for misunderstanding the instructions (2) and for scoring over 2 SDs away from the mean (2).

Stimuli

Experiment 2A. Two familiarization streams were created by time-reversing the two 10-min streams from Experiment 1A every 50 ms. As for the test sounds in Experiment 1, 5-ms attenuation at reversal boundaries prevented noise bursts that come from abrupt changes in amplitude. Test pairs were again composed of the same words and non-words (see **Table 1**). Time-reversed units were identical to the test pairs from Experiment 1B. Non-distorted units were identical to the test pairs used in Experiment 1A.

Experiment 2B. Four familiarization streams were created by time-reversing the four 10-min streams from Experiment 1C every 100 ms. Test pairs were composed of the same words and non-words (see **Table 1**). Units that were time-reversed by 100-ms were identical to the test pairs from Experiment 1C. Non-distorted units were similar to the test pairs used in Experiment 1A, except that 5-ms amplitude attenuations occurred every 100-ms, instead of every 50-ms to parallel the 100-ms time-reversed familiarization stream.

Procedure

The experimental setting, apparatus, and procedure were similar to Experiment 1. In Experiment 2A, participants were familiarized with one of the two 50-ms time-reversed streams. Half of the participants were tested with pairs of 50-ms time-reversed words and non-words, and half with non-distorted pairs of words and non-words. In Experiment 2B, participants were familiarized with one of the four 100-ms time-reversed streams and were either tested with pairs of 100-ms time-reversed words and non-words or pairs of non-distorted words and non-words. As in Experiment 1, participants were asked to choose the more familiar member within each of the 36 pairs.

RESULTS AND DISCUSSION

Since preliminary analyses revealed no difference in performance caused by the presentation of the different randomized time-reversed familiarization streams, the streams were combined in subsequent analyses. In 2A, the proportion of correct answers on the test phase for participants familiarized with 50-ms time-reversed streams was 58.5% on acoustically similar 50-ms time-reversed words, and 57.6% on non-distorted words, both of which were significantly higher than chance [$t(14) = 3.25$, $p = .006$, and $t(14) = 2.81$, $p = .014$, respectively; see **Figure 2**]. These were not different from each other, $F(1, 28) < 1$. For participants familiarized with 100-ms time-reversed streams in 2B, recognition of 100-ms time-reversed words was 55.4% correct, which was significantly above chance [two-tailed one-sample t-test, $t(14) = 3.93$, $p = .002$]. Participants tested on non-distorted words achieved 50.1% correct performance, which was not significantly different from chance [$t(14) = 0.02$, $p =$ ns]. This difference did not reach significance, $F(1, 28) = 2.14$, $p = .15$.

While recognition was high for non-distorted words after 50-ms familiarization, participants did not transfer across acoustic transformations after 100-ms familiarization. These results contrast with recognition performance on distorted words in Experiment 1B and 1C, in which both 50-ms time-reversed and 100-ms time-reversed words were mapped successfully onto representations segmented from the non-distorted familiarization stream. Participants were thus able to recognize highly distorted words after non-distorted familiarization, but were limited in their ability to segment units from highly distorted streams. In this latter case, participants were not able to recognize the units when they were not distorted. This suggests that statistical segmentation processes can function over highly distorted speech input, but the representations that are created are not easily abstracted to non-distorted test items.

EXPERIMENT 3: REPRESENTING STATISTICAL PRODUCTS: ABSTRACTING TO A GENDER VOICE CHANGE

Even under conditions of severe distortion, listeners in Experiment 1 performed better than chance in recognizing segmented units. However in Experiment 2, listeners were limited in their ability to generalize units learned from a distorted stream. In Experiment 3, instead of artificially distorting speech with time-reversals, we test participants on a more natural type of acoustic change, a change in voice (e.g., Church and Schacter, 1994; Goldinger, 1996). The acoustic properties of syllables presented during familiarization and during the test phase are different because they are generated by different voices, but there is no distortion to decrease the intelligibility of the speech sounds. As in Experiment 1, participants in the first condition (3A) were familiarized with a female speech stream, but were tested on a male voice. In the second condition (3B), participants were again familiarized with the female speech stream, but now were tested on a male voice that has been time-reversed with 50-ms time-reversals, a manipulation that had resulted in above chance performance in Experiment 1B.

MATERIALS AND METHODS

Participants

Thirty-two participants were recruited through ads posted on the university campus and were compensated with $5. Participants gave informed consent and reported being fluent in English and having no hearing impairment. Participants were assigned to Experiment 3A (16) or 3B (16). In 3A, three additional participants were tested and excluded from the sample because of experimenter error (1), misunderstanding of the instructions (1) and for being over 2 SDs away from the mean (1). In 3B, three additional participants were tested and excluded from the sample for misunderstanding the instructions (1) and for scoring more than 2 SDs away from the mean (2).

Materials

The two familiarization streams used for the current experiment were identical to the streams presented in Experiments 1A and 1B. For the test trials, the six words and six non-words from Experiment 1 were re-synthesized with a male voice. The synthesis of the male was speech sounds was conducted using DECtalk software in the same manner as in Experiment 1A, with identical parameters, except that the electronic voice "Paul" was used instead of the voice "Betty" and the average pitch was lowered to 100 Hz. For 2B, the male voice words and non-words were locally time-reversed every 50 ms, using the same procedure for time-reversals as in Experiment 1.

Procedure

The experimental setting, apparatus, and procedure were similar to Experiment 1. Participants were familiarized with one of the 10-min female speech streams used in Experiment 1. In the test phase, participants were tested with 36 word and non-word pairs spoken in a male voice that was either normal (3A), or time-reversed (3B). As in Experiment 1, participants were asked to choose the more familiar member of each pair.

RESULTS AND DISCUSSION

When participants were tested on normal male words and non-words (3A), the correct words were selected on average in 59.7% of test trials. A planned one-sample t-test indicated that this performance was significantly above chance [$t(14) = 3.43$, $p = 0.004$; see **Figure 2**]. Performance on items with a voice change was equivalent to testing on normal speech with no voice change in Experiment 1A, $F(1,28) = 1.92$, $p = \text{ns}$, and to testing with mild distortion in Experiment 1B, $F(1,28) < 1$, $p = \text{ns}$. Participants readily mapped newly segmented units onto test items spoken in a different voice (3A), demonstrating that the representation of segmented units is abstract enough to resist voice transformations.

In Experiment 3B with test items that had both a change of voice and 50-ms reversals, the correct words were selected in 50.2% of test trials, a proportion that was no different from chance [$t(14) = 0.1$, $p = \text{ns}$]. Performance was reliably worse than that of participants who were tested on 50-ms reversals of the same voice in Experiment 1, $F(1,28) = 17.9$, $p < 0.001$. Although a time-reversal of 50-ms results in perfectly intelligible speech sounds (as seen in Experiment 1B, and in Saberi and Perrott, 1999), newly segmented units were not recognized when they were spoken in a novel voice that was 50-ms time-reversed. Both types of acoustic changes — voice and time-reversal — on their own have no appreciable effect on performance. However, when these two acoustic changes are combined, performance falls to chance levels. This suggests that the amount of acoustic transformation, and not necessarily the intelligibility of sounds, can influence the recognition of putative lexical units newly segmented through statistical processes. These units are thus not represented in fully abstract terms.

GENERAL DISCUSSION

In three experiments we examined the nature of representations generated by statistical segmentation processes. Learners were able to recognize segmented units when trained with non-distorted speech and tested on mild and high distortions, when trained on mildly or highly distorted speech and tested on the same distorted speech, when trained on mild distortions and tested on non-distorted speech, and when trained on a female voice and tested on a male voice. In contrast, recognition was not successful when learners were trained on highly distorted speech and tested on non-distorted speech and when learners were trained on a non-distorted female voice and tested on a male voice with mild distortion.

Adults' success in generalizing segmented units to acoustically distinct units suggests that the products of segmentation are not stored entirely as acoustically based stimulus-bound representations (as in Conway and Christiansen, 2006), nor are they fully abstract (as in Turk-Browne et al., 2005). Adults' uniform success in segmenting words with mild distortion even when the testing mode differed acoustically from the familiarization mode suggests both that learners can compensate for distortions in the speech stream during acquisition, and that the segmented units are partly stored as abstract representations. Conversely, familiarization on highly distorted speech leads to the acquisition of acoustic representations of the units, since learners cannot generalize to normal words.

Learning patterns with the current materials were likely influenced by the exposure time and the consolidation time of the experimental paradigm. As in previous statistical segmentation studies (Peña et al., 2002; Toro et al., 2008), adults were exposed to the familiarization streams for 10 min and were tested immediately after training on a two-alternative forced choice test. This training and test regime resulted in a pattern of learning in which learners were able to generalize across some distortions (e.g., training on non-distorted speech and testing on high distortions) but not others (e.g., training on highly distorted speech and testing on non-distorted speech). Much like consolidation timing can influence the integration of spoken words into the lexicon (Gaskell and Dumay, 2003; Dumay and Gaskell, 2007) and longer interstimulus intervals can facilitate phonemic — rather than acoustic — encoding (Pisoni, 1973; Werker and Tees, 1984), a longer consolidation time could have allowed learners to create more abstract representations and permitted broader generalizations. Given identical training and testing parameters, however, differences in learning were observed across the different experiments.

Words with 100-ms reversals can be acquired but not generalized across acoustic contexts, contrary to non-distorted speech and 50-ms time-reversed speech. The asymmetry between acquiring from and generalizing to highly distorted speech suggests that the processes involved in acquisition are not necessarily the same as are involved in generalization (Marcus et al., 2007). It appears easier to recover already segmented units when they are distorted, than to learn from highly distorted units, suggesting that acquisition processes might be more vulnerable to variability.

What accounts for the failure to generalize from highly distorted speech? Recall, participants showed evidence of segmenting the highly distorted speech stream when they were tested on highly distorted units, but they did not generalize to normal speech units. One possibility is that the impairment on generalization from 100-ms time-reversals to normal speech is a mere function of the acoustic difference between the two stimuli. The success of segmentation might depend on the degree of acoustic similarity between the familiarization stream and test items. This is unlikely to account for the failure because learners were able to generalize across an acoustic difference of equal magnitude in Experiment 1 when they learned from normal speech and generalized to 100-ms time-reversals.

A second possibility is that the creation of an abstract representation is more effortful than simply encoding an exact acoustic representation of the stimuli. Distortion in the speech stream may increase the processing demands of the task of statistical learning; it may be easier to pay attention to intelligible speech than to highly distorted speech, or the fact that such distorted speech is a stimulus that listeners are not familiar with may lead to higher processing requirements (Davis et al., 2005). Increased attention demands for processing the distortion may leave less resources available for the creation of a more abstract representation, if indeed the creation of an abstract representation requires more processing resources. Though previous studies demonstrated that statistical learning could occur both in adults and children while they are engaged in an unrelated activity (Saffran et al., 1997), more recent research has shown that, when participants are engaged in a task that is very attention-demanding or when their attention is directed to irrelevant aspects of the material, statistical learning sometimes fails (Baker et al., 2004; Toro et al., 2005; Turk-Browne et al., 2005).

A third possibility is that generalizing to an acoustically dissimilar stimulus could depend on the intelligibility of the stimulus available during acquisition. The acquisition of an abstract representation might necessitate the presence of recognizable phonetic information in the input. The sounds constituting the normal speech in the present experiment were all phonemes that are present in the English language. Participants may have been mapping the sounds they heard onto already existing abstract phonetic patterns in their language, thereby facilitating generalization of normal speech syllables to distorted sounds. Fifty-millisecond time-reversed speech might still sufficiently sound like English for this mapping process to occur, but 100-ms time-reversed speech might just be too distorted for mapping onto phonemes to occur during acquisition. Even known words that are time-reversed by 100 ms cause a 50% drop in participants' speech recovery rates (Saberi and Perrott, 1999).

More broadly, generalization may depend on the different effect of time-reversals on different segments of speech. Consonants have been proposed to play an important role in lexical processing, while vowels play a role in indexical, prosodic, and grammatical processing (Nespor et al., 2003; Bonatti et al., 2007). Empirical findings provide support for differences between processing of consonants and vowels: adults extract statistical regularities from consonants but not vowels (Bonatti et al., 2005; Toro et al., 2008; but see Newport and Aslin, 2004), while rule-extraction is privileged over vowels but not consonants for adults and infants (Bonatti et al., 2005; Toro et al., 2008; Pons and Toro, 2010; Hochmann et al., 2011). The privileged role of consonants in statistically based segmentation computations may account for the more limited generalization of segments acquired from highly distorted speech. Because the 100-ms time-reversal time window was of a longer duration than most consonants, consonants were more distorted than in the 50-ms time-reversed condition. Perhaps the absence of intelligible consonant sequences in the highly distorted stream prevented any encoding of the acquired units as potential lexical forms. Since 50-ms-time-reversed speech is close to perfectly intelligible (Saberi and Perrott, 1999), the consonants may have still been perceptible in these mildly distorted streams. However, consonant intelligibility may not fully account for the failure to generalize across all experiments, since the slightly distorted male words and non-words that were tested in Experiment 3 were perfectly intelligible, yet learners were still unable to recognize units segmented from a non-distorted female stream.

What may be essential for generalization is for listeners to have pre-existing categorical representations, whether linguistic or not, on which they are able to map the input. Speech sounds may be represented as categories centered on prototypical instances (e.g., Rosch, 1975; Samuel, 1982). Prototypical stop consonants have been shown to lead to priming of both prototypical and non-prototypical instances of that consonant, while non-prototypical instances only prime identical non-prototypical instances (Ju and Luce, 2006). Even infants generalize more readily to novel instances after being familiarized with a prototypical vowel compared with non-prototypical exemplars (Grieser and Kuhl, 1989). This suggests that mapping onto a prototypical existing representation might indeed facilitate generalization. The distorted speech sounds could be encoded as less prototypical speech sounds and would therefore be generalized less.

The representations formed by learners were abstracted across some, but not all, acoustic contexts. This opens two possibilities for how products of statistical learning might contribute to word learning. One possibility is that units segmented through purely statistical computations play very little role in word learning. Consistent with this view, learners compute and extract associations between syllables without encoding words unless additional cues are present (as proposed by Endress and Mehler, 2009). Alternatively, even fragile units segmented by statistical processes may function as words in elementary word learning tasks (Graf Estes et al., 2007; Mirman et al., 2008). In natural word learning, it may be that additional mechanisms reify representations that are extracted by statistical processes during segmentation.

CONCLUSION

This study explored the type of representations that are created for units acquired through statistical segmentation processes. We demonstrate that statistical learning of acoustic units is possible following a 10-min exposure to speech that is non-distorted, mildly distorted (with 50-ms time-reversals), and highly distorted (with 100-ms time-reversals). The representations formed by learners can be abstracted across some acoustic contexts, but not others. We propose that learners map newly learned units onto existing phonetic representations when they can, but the failure to recognize intelligible units in a different voice that was mildly distorted suggests that segmented units may not function as full lexical forms.

ACKNOWLEDGMENTS

This research was funded by grants from the Natural Sciences and Engineering Research Council of Canada and NYU research funds to Athena Vouloumanos.

REFERENCES

Altmann, G. T. M., Dienes, Z., and Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 899–912.

Assmann, P., and Summerfield, Q. (2004). The perception of speech under adverse conditions. *Springer Handbook of Auditory Research* 18, 231–308.

Baker, C. I., Olson, C. R., and Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychol. Sci.* 15, 460.

Blesser, B. (1972). Speech perception under conditions of spectral transformation. I. Phonetic characteristics. *J. Speech Hear. Res.* 15, 5–41.

Bonatti, L. L., Peña, M., Nespor, M., and Mehler, J. (2005). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychol. Sci.* 16, 451–459.

Bonatti, L. L., Peña, M., Nespor, M., and Mehler, J. (2007). On consonants, vowels, chickens, and eggs. *Psychol. Sci.* 18, 924–925.

Church, B. A., and Fisher, C. (1998). Long-term auditory word priming in preschoolers: implicit memory support for language acquisition. *J. Mem. Lang.* 39, 523–542.

Church, B. A., and Schacter, D. L. (1994). Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency. *J. Exp. Psychol. Learn.* 20, 521–533.

Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behav. Res. Meth.* 25, 257–271.

Cole, R. A., Jakimik, J., and Cooper, W. E. (1980). Segmenting speech into words. *J. Acoust. Soc. Am.* 67, 1323–1332.

Conway, C. M., and Christiansen, M. H. (2006). Statistical learning within and between modalities: pitting abstract against stimulus-specific representations. *Psychol. Sci.* 17, 905–912.

Curtin, S., Mintz, T. H., and Byrd, D. (2001). "Coarticulatory cues enhance infants' recognition of syllable sequences in speech," in *Proceedings of the 25th Annual Boston University Conference on Language Development* (Somerville: Cascadilla Press), 190–201.

Davis, M. H., and Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hearing Res.* 229, 132–147.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *J. Exp. Psychol. Gen.* 134, 222–241.

Dumay, N., and Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychol. Sci.* 18, 35–39.

Endress, A. D., and Mehler, J. (2009). The surprising power of statistical learning: when fragment knowledge leads to false memories of unheard words. *J. Mem. Lang.* 60, 351–367.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behav. Dev.* 8, 181–195.

Fiser, J., and Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15822–15826.

Gaskell, M. G., and Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition* 89, 105–132.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279.

Goldinger, S. D., Pisoni, D. B., and Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *J. Exp. Psychol. Learn.* 17, 152–162.

Graf Estes, K., Evans, J. L., Alibali, M. W., and Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol. Sci.* 18, 254–260.

Greenberg, S., and Arai, T. (2004). What are the essential cues for understanding spoken language? *IEICE Trans. Inf. Syst.* 87, 1059–1070.

Grieser, D. A., and Kuhl, P. K. (1989). Categorization of speech by infants: support for speech-sound prototypes. *Dev. Psychol.* 25, 577–588.

Hauser, M. D., Newport, E. L., and Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* 78, B53–B64.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., and Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: effects of feedback and lexicality. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 460–474.

Hochmann, J. R., Benavides-Varela, S., Nespor, M., and Mehler, J. (2011). Consonants and vowels: different roles in early language acquisition. *Dev. Sci.* 14, 1445–1458.

Johnson, E. K., and Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: when speech cues count more than statistics. *J. Mem. Lang.* 44, 548–567.

Ju, M., and Luce, P. A. (2006). Representational specificity of within-category phonetic variation in the long-term mental lexicon. *J. Exp. Psychol. Human* 32, 120–138.

Jusczyk, P. W., Hohne, E. A., and Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Percept. Psychophys.* 61, 1465–1476.

Jusczyk, P. W., and Luce, P. A. (2002). Speech perception and spoken word recognition: past and present. *Ear Hear.* 23, 2–40.

Jusczyk, P. W., Pisoni, D. B., and Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition* 43, 253–291.

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 83, B35–B42.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *J. Phon.* 3, 129–140.

Kraljic, T., and Samuel, A. G. (2005). Perceptual learning for speech: is there a return to normal? *Cogn. Psychol.* 51, 141–178.

Liu, S. A. (1996). Landmark detection for distinctive feature-based speech recognition. *J. Acoust. Soc. Am.* 100, 3417–3430.

Marcus, G. F., Fernandes, K. J., and Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychol. Sci.* 18, 387–391.

Mattys, S. L. J., and Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78, 91–121.

Miller, G. A., and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* 27, 338.

Mirman, D., Magnuson, J. S., Graf Estes, K., and Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition* 108, 271–280.

Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *J. Acoust. Soc. Am.* 85, 365–378.

Nespor, M., Peña, M., and Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio* 2, 221–247.

Newport, E. L., and Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychol.* 48, 127–162.

Newport, E. L., Hauser, M. D., Spaepen, G., and Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cogn. Psychol.* 49, 85–117.

Peña, M., Bonatti, L. L., Nespor, M., and Mehler, J. (2002). Signal-driven computations in speech processing. *Science* 298, 604–607.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Atten. Percept. Psychophys.* 13, 253–260.

Pons, F., and Toro, J. M. (2010). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition* 116, 361–367.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science* 212, 947–949.

Rosch, E. (1975). Cognitive representations of semantic categories. *J. Exp. Psychol. Gen.* 104, 192–233.

Saberi, K., and Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature* 398, 760.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996b). Word segmentation: the role of distributional cues. *J. Mem. Lang.* 35, 606–621.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition* 70, 27–52.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., and Barrueco,

S. (1997). Incidental language learning: listening (and learning) out of the corner of your ear. *Psychol. Sci.* 8, 101–105.

Samuel, A. G. (1982). Phonetic prototypes. *Atten. Percept. Psychophys.* 31, 307–314.

Samuel, A. G. (1988). Central and peripheral representation of whispered and voiced speech. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 379–388.

Scott, S. K., and Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* 26, 100–107.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304.

Stevens, K. N. (1980). Acoustic correlates of some phonetic categories. *J. Acoust. Soc. Am.* 68, 836–842.

Stickney, G. S., and Assmann, P. F. (2001). Acoustic and linguistic factors in the perception of bandpass-filtered speech. *J. Acoust. Soc. Am.* 109, 1157–1165.

Swingley, D., and Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition* 76, 147–166.

Swingley, D., and Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychol. Sci.* 13, 480–484.

Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy* 7, 53–71.

Thiessen, E. D., and Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Dev. Psychol.* 39, 706–716.

Tillmann, B., and McAdams, S. (2004). Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)similarities. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1131–1142.

Toro, J. M., Nespor, M., Mehler, J., and Bonatti, L. L. (2008). Finding words and rules in a speech stream: functional differences between vowels and consonants. *Psychol. Sci.* 19, 137–144.

Toro, J. M., Sinnett, S., and Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition* 97, B25–B34.

Toro, J. M., and Trobalon, J. B. (2005). Statistical computations over a speech stream in a rodent. *Percept. Psychophys.* 67, 867–875.

Turk-Browne, N. B., Junge, J., and Scholl, B. J. (2005). The automaticity of visual statistical learning. *J. Exp. Psychol. Gen.* 134, 552–564.

Van Tasell, D. J., Greenfield, D. G., Logemann, J. J., and Nelson, D. A. (1992). Temporal cues for consonant recognition: training, talker generalization, and use in evaluation of cochlear implants. *J. Acoust. Soc. Am.* 92, 1247–1257.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science* 167, 392–393.

Werker, J. F., and Curtin, S. (2005). PRIMIR: a developmental framework of infant speech processing. *Lang. Learn. Dev.* 1, 197–234.

Werker, J. F., and Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *J. Acoust. Soc. Am.* 75, 1866–1878.

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 October 2011; accepted: 23 February 2012; published online: 23 March 2012.

Citation: Vouloumanos A, Brosseau-Liard PE, Balaban E and Hager AD (2012) Are the products of statistical learning abstract or stimulus-specific? Front. Psychology ***3****:70. doi: 10.3389/fpsyg.2012.00070*

This article was submitted to Frontiers in Language Sciences, a specialty of Frontiers in Psychology.