Phonological discrimination and contrast detection in pupillometry

Julia S. C. Chiossi[1,2]*, François Patou[3], Elaine Hoi Ning Ng[1,4], Kathleen F. Faulkner[1] and Björn Lyxell[2]

[1]Oticon A/S, Smørum, Denmark, [2]Department of Special Needs Education, University of Oslo, Oslo, Norway, [3]Oticon Medical, Smørum, Denmark, [4]Department of Behavioural Sciences and Learning, Linnaeus Centre HEAD, Swedish Institute for Disability Research, Linköping University, Linköping, Sweden

Introduction: The perception of phonemes is guided by both low-level acoustic cues and high-level linguistic context. However, differentiating between these two types of processing can be challenging. In this study, we explore the utility of pupillometry as a tool to investigate both low- and high-level processing of phonological stimuli, with a particular focus on its ability to capture novelty detection and cognitive processing during speech perception.

Methods: Pupillometric traces were recorded from a sample of 22 Danish-speaking adults, with self-reported normal hearing, while performing two phonological-contrast perception tasks: a nonword discrimination task, which included minimal-pair combinations specific to the Danish language, and a nonword detection task involving the detection of phonologically modified words within sentences. The study explored the perception of contrasts in both unprocessed speech and degraded speech input, processed with a vocoder.

Results: No difference in peak pupil dilation was observed when the contrast occurred between two isolated nonwords in the nonword discrimination task. For unprocessed speech, higher peak pupil dilations were measured when phonologically modified words were detected within a sentence compared to sentences without the nonwords. For vocoded speech, higher peak pupil dilation was observed for sentence stimuli, but not for the isolated nonwords, although performance decreased similarly for both tasks.

Conclusion: Our findings demonstrate the complexity of pupil dynamics in the presence of acoustic and phonological manipulation. Pupil responses seemed to reflect higher-level cognitive and lexical processing related to phonological perception rather than low-level perception of acoustic cues. However, the incorporation of multiple talkers in the stimuli, coupled with the relatively low task complexity, may have affected the pupil dilation.

KEYWORDS

pupillometry, speech perception, phoneme perception, acoustic cues, novelty detection, linguistic context

1. Introduction

The perception of contrast between phonemes is a fundamental aspect of speech perception and the basis for language acquisition (Kuhl et al., 2008; Casserly and Pisoni, 2010). By gradually extracting patterns from speech, infants learn to divide acoustic input into phonetic categories and sequences of phoneme combinations into words (Kuhl et al., 2008; Romberg and Saffran,

2010). As vocabulary grows, consistent perception of acoustic/phonetic patterns in the speech input will activate lexical processing for word recognition or pathways for learning a novel word (Pittman et al., 2017). However, the perception of phonological contrasts is not uniquely driven by its acoustic properties. Phonological perception evolves to accommodate predictions from the linguistic context and the inherent phonetic variability in speech (Repp, 1982; Allen et al., 2003; Clarke and Garrett, 2004; Jesse, 2021). In terms of cognitive processing, low-level acoustic perception and high-level lexical processing are integrated to determine the presence or significance of a specific contrast (Borsky et al., 1998; Coleman, 2003).

Low-level processing involves interpreting the acoustic properties of speech sounds. Acoustic cues refer to distinct auditory features that convey information. Phoneme contrasts are marked by a variety of cues in spectral and temporal acoustic dimensions. These acoustic cues are redundant, such that several distinct cues occur for a particular contrast and can be traded with each other for the perception of a particular phoneme (Liberman et al., 1967; Repp, 1982; Winn et al., 2012). As an example, place of articulation for stop consonants can be cued by F2 and F3 transitions, burst frequency, or burst amplitude (Repp, 1982, for multiple examples). These cues covary in natural speech, and listeners must integrate them to achieve the most reliable identification of the incoming speech stimuli. Although this process may remain robust while a cue is missing or degraded, it would be expected that the demand for cognitive processing would increase with higher ambiguity of the stimulus and degradation of differential cues.

High-level phonological processing is guided by top-down knowledge of linguistic rules and context (Casserly and Pisoni, 2010; Jesse, 2021). The presence of high-level processing on phoneme discrimination leads to a perceptual bias in which listeners disambiguate underspecified phonemes toward meaningful compositions (Ganong, 1980; Jesse, 2021). For example, individuals may identify ambiguous speech sounds as a real word rather than a nonsense word if context is present.

For this high-level component in phonological perception, it is challenging to separate the specific contribution of each type of processing on task accuracy, as both high-level and low-level processing may be involved in successful performance. However, an ambiguous stimulus, which theoretically requires higher levels of processing, would likely demand the allocation of more cognitive resources, thus increasing listening effort (Kramer et al., 2013; Johnsrude and Rodd, 2016). Therefore, having an objective method that is sensitive to the individual demands for both low and high-level phonological processing could help to explain the variability in speech perception performance attributed to hearing impairment and challenging auditory environments (Phatak and Grant, 2014; Gianakas and Winn, 2019).

Assessing cognitive processes objectively during phonological contrast perception involves measuring responses to speech stimuli occurring at a cortical level. Physiological markers like late-latency auditory event-related potentials and mismatch negativity (MMN) have provided evidence of a pre-attentive component of phonological discrimination which could capture changes in the phonological pattern before conscious perception (Näätänen et al., 2007; Steinhauer and Connolly, 2008). However, cortical measures can be time-consuming and uncomfortable for participants. As an alternative, pupillometry could be used as a tool to investigate the temporal dynamics between low- and high-level processing of phonological stimuli. The pupil dilation is linked to increasing the norepinephrine release from the locus coeruleus, an area associated with attentional prioritization and perception of novelties (Eckstein et al., 2017; Kafkas and Montaldi, 2018). Task-evoked pupil dilation is measured by videorecording the pupil size during a task. It has the advantages of being cost-effective and providing good temporal resolution (Winn et al., 2018). In terms of cognitive processing, at a low level, pupil dilation is sensitive to the perception of novelty that arises from a mismatch between stimulus and context, as changes in the stimulus frequency, intensity or pitch (Virtala et al., 2018; Bala et al., 2020). At a high level, pupil dilation has been shown to increase with linguistic processing demand, working memory load, and the effort required to resolve ambiguity in speech (Wendt et al., 2016; Zekveld et al., 2018; Kadem et al., 2020; Micula et al., 2022).

Previous research on the use of pupillometry to assess the discrimination of speech sounds, particularly phonemes, is limited and mainly focused on perception within word or sentence context (Wagner et al., 2016; Kinzuka et al., 2020; Winn and Teece, 2021). Kinzuka et al. (2020) used an oddball paradigm to evaluate the correlation between the perceptual ability of Japanese speakers to discriminate the /r/ and /l/ sounds in real words and their pupillometric responses. The study found that higher language proficiency is associated with earlier occurring differences in peak of pupil dilation (PPD) between target frequent and infrequent stimuli. In another study addressing the effects of phonological manipulation on pupil dynamics, Winn and Teece (2021, 2022) measured pupil dilation during sentence recognition using embedded phonologically altered words in both normal hearing and cochlear implant subjects. The authors reported a steeper increase in pupil size in response to phonological alterations, with larger differences when the target phoneme was substituted by noise instead of another phoneme. For cochlear implant users, the contrast in pupil responses between sentences with and without phonological substitutions was shallower than for normal hearing peers, suggesting a relationship between degraded speech perception and the pupil response. This effect of speech degradation was also observed by Wagner et al. (2016), who reported steeper pupil dilation curve slopes for unexpected word prosody in full-frequency-spectrum speech but not for cochlear implant simulated speech.

It is important to note that in the paradigms described above, participants' attention was directed toward processing the entire sentence or word, actively engaging high-level processing to interpret its meaning. However, the presence of context makes it difficult to distinguish the pupil variation caused by the perception of a phonological contrast, from that caused by the perception of lexical meaning variation which is known to cause pupil dilation independently (Kamp and Donchin, 2015). Additionally, sentence comprehension in adverse listening conditions is known to produce higher pupil dilation (Wendt et al., 2016; Ohlenforst et al., 2018; Trau-Margalit et al., 2023). Therefore, directing attention to the whole sentence might obscure the responses related to phonological contrast detection, since pupillometry seems to reflect responses to attended stimuli rather than passive listening (Kramer et al., 2013) and is larger for perceived rather than unperceived errors (Kamp and Donchin, 2015). On the contrary, it is possible that by directing attention to the presence of phonologically altered stimuli, responses may better reflect the low-level processing independently of the presence of context.

This study aimed to investigate the pupil temporal dynamics during the auditory processing of phonological contrasts, in an effort to differentiate low- and high-level processing of phonological information. For that, two paradigms were contrasted. First, we investigated the pupillometric response to low-level processing of phonological contrasts, measured during the perception of lexically decontextualized phoneme contrasts (phonological discrimination task). Second, we explored the possibility to record similar responses in the presence of lexical information, without prompting high-level sentence processing (detection task). Additionally, to introduce an acoustic challenge to the perception of the phonological contrast, we investigate how those responses are impacted a sub-optimal speech input, using a vocoded speech signal.

To minimize the influence of lexical knowledge on phonological perception, we chose to explore phonological contrasts using nonwords as our tokens. This approach aimed to preserve the real-world relevance of word-like items, while removing their lexical meaning. Traditional methods for assessing phonological identification and recognition often employ syllabic continua (Iverson, 2003; Abada et al., 2008; Lewis and Bidelman, 2020). However, syllables may not demand the same level of cognitive processing for their phonological contrasts as longer stimuli. Therefore, we chose to use nonwords, as they would enhance the ecological validity of the pupillometry measures.

We hypothesized that if pupil dynamics were sensitive to the low-level acoustic properties of a phonological contrast, larger pupil dilations would be measured in conditions where the phonological contrast is present. These differences would be maintained independently of the presence of context and under a vocoded speech signal, whenever the phonological contrast was correctly perceived. However, if the pupil dynamics reflect the high-level linguistic processing required to disambiguate a phonological contrast, larger pupil dilations would be measured for a phonological contrast only in the presence of lexical information. We expect that the results provided here enhance the understanding in pupillary responses to phonological processing, shedding light on their sensitivity to different processing levels and adverse speech conditions.

2. Materials and methods

2.1. Participants

A convenient sample was recruited across the researcher's place of employment, in the Capital Region of Denmark. A sample of 22 adults (age: [25; 0–65;0], median: 50 years; females: 39%; all with more than 11 years of education) who reported Danish as their first language, and self-reported normal hearing, were included. Participation was voluntary, and the researchers were contacted directly by the participants after an online announcement in an internal website. Participants joined during working hours and were compensated with their regular salary for their time.

This study was waived from ethical review by the Regional Committee of Health Research Ethics - Capital Region, Denmark, after inquiry submission, as it was considered to be research in the social domain. All participants gave active consent to the study, after receiving written and oral information, in accordance with the Declaration of Helsinki. Requirements regarding the General Data Protection Regulation (GDPR) were carefully followed.

2.2. Stimuli

2.2.1. Phonological discrimination task

The discrimination task was composed of two lists containing 40 pairs of disyllabic-nonwords, selected from the nonword corpus published by Nielsen and Dau (2019). From the original material, C−/a/-C−/a/ nonwords starting with one of the 14 main initial phonemes for the Danish language (/p t k b d g m n l f v s r h /) were selected. The recordings selected had 100% speech intelligibility score reported in the original study by Nielsen and Dau (2019), for both first and second consonants. The present study focuses only on the first phoneme contrast.

For the task, nonwords were combined in pairs, to account for all the minimal-pair combinations in Danish for which just one distinctive production feature is present (place, voice/aspiration, or manner, as in 'bafi − 'pafi'). To facilitate phonological ─ instead of acoustical ─ comparison of the word pairs, recordings of each nonword from three different speakers were selected from the original material and each pair was presented using audio from two different speakers, selected randomly among the three possible recordings. Post-hoc analysis revealed no effect of the speaker-pair used on participants' performance.

All the audio recordings were normalized by root mean square (RMS) and silence was added in the beginning of each file to align the nonwords' offset during the task and to randomize the nonword start and the interval between nonwords.

2.2.2. Detection task

A second task explored the effect of context, in which the participants were asked to track a phoneme substitution in a word within a sentence (Pittman and Schuett, 2013). The detection task was composed of two lists of 36 four-word-sentences. The original sentences in the lists were composed of simple words from a 3 year-old child's vocabulary (Bleses et al., 2008a, 2008b) to guarantee that the words included would be well known by the participants. The sentences were evaluated as highly meaningful by a group of 25 native Danish speakers, in a pre-study conducted by our group. For half of the sentences in each list, the first phoneme of the second word was substituted for another phoneme with similar phonotactic probability (e.g., *Hunden finder altid maden* [the dog always finds the food] ->*Hunden* sinder *altid maden*, an equivalent example in English would be 'Dad buys new shirts' - > 'Dad fuys new shirts', from Pittman and Schuett (2013)). We were careful to choose phonemes that would generate a nonword when replacing the original phoneme, which was confirmed by a group of 14 native speakers who listened to the generated nonwords in isolation and were asked to write the first real word it would remind them of (less than 50% of the participants could point to same original or other real word). Due to this requirement, the phonemes selected had contrasts in one or more production features with the original phoneme, which potentially added cues that may have aided detection in the sentence context. Moreover, to avoid that the second word in the sentence could be predicted by the sentence context, the same group of 25 native speakers were asked to complete the sentences where the target word was missing. Only the sentences with less than 10% of participants filling in the same real word (defined as low cloze probability in Kutas and Hillyard, 1984) were included in the lists.

The final 72 sentences, half with embedded nonwords, were recorded by a female native Danish speaker with an accent from the

Danish capital region. She was instructed to pronounce the sentences in a natural prosody but in a slow speaking pace. All the recordings were normalized by RMS and silence was added in the beginning of each file to randomize the sentences' start.

2.2.3. Stimuli vocoding

In order to reduce the acoustic features, challenging the detection and discrimination of phonological contrasts (Stenfelt and Rönnberg, 2009), one list was randomly vocoded for each participant. The vocoding process includes dividing the speech signal into frequency bands, extracting the amplitude envelope for each band, and using it to modulate a noise band, resynthesizing the bands to create a new audio file. For this study, the vocoded versions of the stimuli were generated using the software Praat (Boersma and Weenink, 1992) and the open-source code provided by Winn (2021) (version 45). An 8-channel vocoder was used, with flat-spectrum noise-carrier, and corner frequencies set between 0.2-8 kHz. This number of bands was chosen to add challenges to the transmission of spectral information while approaching the asymptotic speech-recognition performance in quiet (Dorman et al., 1997; Friesen et al., 2001; Xu et al., 2005).

2.2.4. Vocoded real-word recognition

Considering that a participant's inability to recognize real words in the vocoded condition could influence their performance on nonword detection, vocoded word recognition scores were also calculated. The first list of the clinical test Dantale I (Elberling et al., 1989) in silence was vocoded using the method described above. Participants were presented with the recorded monosyllabic words in isolation and were asked to repeat them aloud. The participant's response was recorded and transcribed, for offline scoring.

2.3. Pupillometry

Pupil size was continually measured by the Pupil Core® platform (Pupil Labs GmbH, Berlin). The glasses-mounted solution includes one front camera recording the gaze direction and two infra-red cameras that record the pupils at a sampling frequency of 200 Hz. Pupil tracking is done in dark mode. The software provides the pupil size for each eye in arbitrary units (pixels) and a confidence score, defined as an index, between 0 and 1, indicating the quality of the acquired value.

2.4. Procedure

The study protocol was implemented via computer on the OpenSesame platform (Mathôt et al., 2012), using the features developed by Sulas et al. (2022).

The experiment was conducted in an acoustically treated sound studio. Participants indicated their responses using a touchscreen monitor placed on a table in front of them. The monitor was positioned to have the top ¾ of the screen aligned to the participant's eye. Sound was presented from a loudspeaker positioned 1 m directly in front of the participant (0-degrees Azimuth). Test participants wore the pupillometry glasses with the cameras adjusted so that the pupils were in the middle of the cameras respective field of view. The glasses were worn during the whole session and adjusted as needed between

tasks in case of displacement. Lighting conditions and the screen luminance were kept constant at 200 lumens.

Prior to starting the experimental tasks, the participants were familiarized with vocoded speech. Sentences were presented back-to-back in non-vocoded and vocoded conditions, for about 3 min, until the participant reported feeling comfortable recognizing the sentence in the vocoded version. The full testing session included other speech perception tasks not reported here and took approximately 1.5 h. Task order and sequence were randomized to counterbalance fatigue effects. All tests were preceded by verbal and written instructions, plus a training phase during which direct verbal feedback and clarifications were provided.

In the phonological discrimination task, word pairs were presented one by one. The participant was asked to indicate if the second word in the pair was the same as the first in a 'yes/no' paradigm. A fixation dot was kept in the screen from 2 s before until 2 s after the presentation of each word pair (detailed in Figure 1). Participants were asked to look at the dot in order to reduce eye movements and improve the quality of the pupillometry data. After each pair presentation, participants indicated their response via touchscreen. This task took approximately 7 min to complete in each condition, and conditions were randomized across participants.
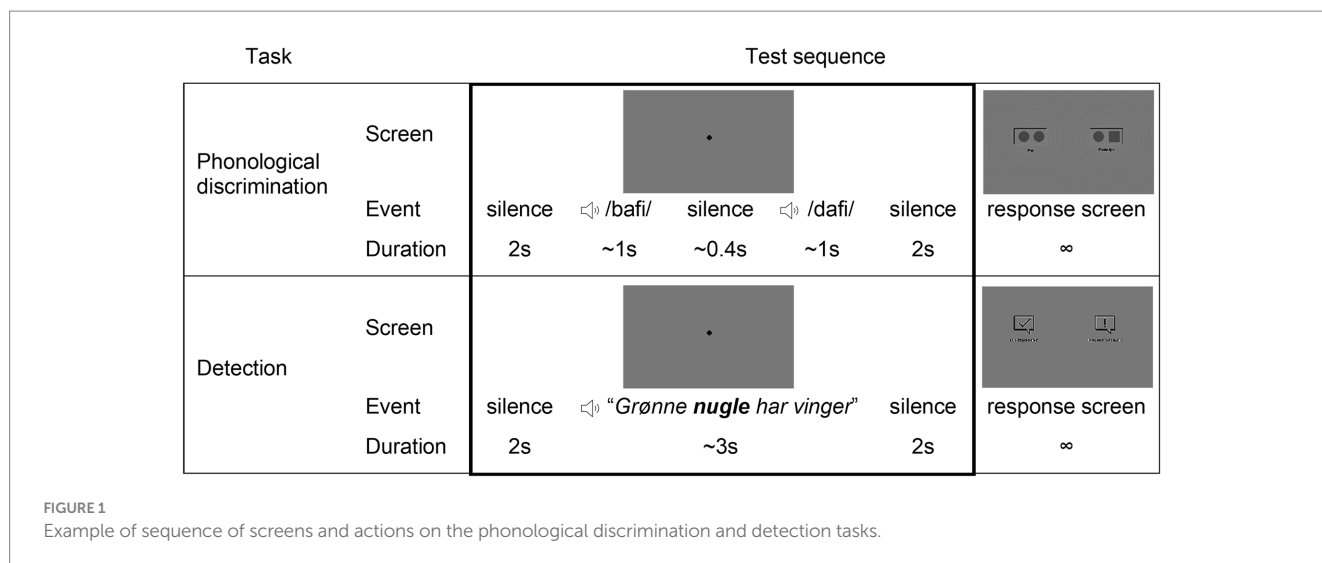
In the detection task, participants were asked to indicate if the sentence contained a nonword (the phonologically modified word) in a 'yes/no' paradigm. Participants listened to a list of 36 sentences in each condition. The trial sequence is illustrated in Figure 1. Two seconds of silence were added before and after each sentence, while a fixation dot was kept on the screen. Testing took approximately 5 min in each condition, and conditions were randomized across participants.

2.5. Analysis

2.5.1. Task performance

For both phonological discrimination and the detection tasks, accuracy for 'yes/no' responses was recorded. Analysis was conducted in terms of signal detection theory (Macmillan and Creelman, 2005). The 'signal' in the stimulus was defined as the presence of a phonological contrast, namely, the presence of a nonword in the sentence or a phonological substitution in the second token of the nonword-pair. Responses were classified as 'hits': correct responses when the signal was present, 'misses': incorrect responses when the signal was present, 'correct rejections': correct responses when the signal was absent, and 'false alarms': incorrectly reporting the presence of the signal when it was absent. The proportion of correct responses was calculated as the sum of 'hits' and 'correct rejections', divided by the total number of trials (Macmillan and Creelman, 2005).

The discrimination score (d') was calculated as measure of the participants' sensitivity to the presence of a signal (Macmillan and Creelman, 2005). It was estimated by subtracting the z-transformed 'hit' rates and 'false-alarm' rates. To avoid floor and ceiling effects in d' calculation, a correction for the extreme values was performed using the log linear approach described by Stanislaw and Todorov (1999), by adding 0.5 to both the number of 'hits' and 'false alarms' and adding 1 to the number of trials, before calculating the d' score. Additionally, to analyze a possible response bias toward selecting one of the two options ('yes'/'no'), the criterion location was calculated as

FIGURE 1

Example of sequence of screens and actions on the phonological discrimination and detection tasks.

minus half of the sum of z-transformed 'hit' and 'false-alarms'. A positive criterion value indicates a bias to 'miss' the signal although it is present, while negative values represent bias toward accusing the presence of the signal despite its absence ('false alarms'). Together, d' and criterion location give a parameter of participants' strategy in the phonological discrimination and detection tasks.

2.5.2. Pupillometry pre-processing and analysis.

Pupil data were segmented by trial, and data were analyzed from the eye with best overall confidence during the task, calculated by the percentage of data points over 0.85 of confidence, as reported by the equipment software. The data were cleaned of blinks and artifacts by detecting dilation speed outliers with the method described by Kret and Sjak-Shie (2019) and excluding the flagged data points with a backward and forward margin of 50 ms. Data reconstruction was done using Piecewise Cubic Hermite Interpolating Polynomial (Pchip) or linear interpolation when the Pchip was not possible (where there were not enough points available before or after the region to interpolate), considered the good reconstruction properties of both methods reported by Dan et al. (2020). Blinks above 500 ms were not reconstructed. The individual data points were downsampled to 30 Hz, as the pupil response latency of is over 200 ms (Winn et al., 2018; Mathôt and Vilotijević, 2022), and smoothed using a moving-average filter of 0.1 s.

Trials with more than 45% interpolated data were excluded from the analysis (Burg et al., 2021; Zhang et al., 2022). Baseline pupil size was calculated per-trial by taking the mean pupil size during the 500 ms right before stimulus onset (Seropian et al., 2022). All subsequent data points in the trial were calculated as the proportional change relative to that baseline pupil size. As a last step, raw and processed data were visually inspected to identify and exclude trials with potential contamination, as artifacts in the baseline estimation period or absolute changes in pupil size over 40% of the baseline (Winn, 2016; Winn et al., 2018).

Subjects with more than 50% of the trials excluded from one task condition, had their results excluded from the analysis in that specific task. This criterion excluded pupillometric data from three subjects in both conditions of the phonological discrimination task only. For the remaining participants and tests, the aggregated trace of the pupil

response for correct answers was calculated. Data was extracted regarding the value and time of the maximum pupil size – respectively, peak pupil dilation (PPD) and the peak pupil dilation latency (PPL) – from the time window spamming from the target stimulus onset (the second word) to 1 s after the audio offset. To compare with studies with similar methodology (Wagner et al., 2016; Winn and Teece, 2021), a growth curve analysis (GCA) was carried out, which models the quadratic fit of the pupil curve between the target-stimulus onset and the PPD.
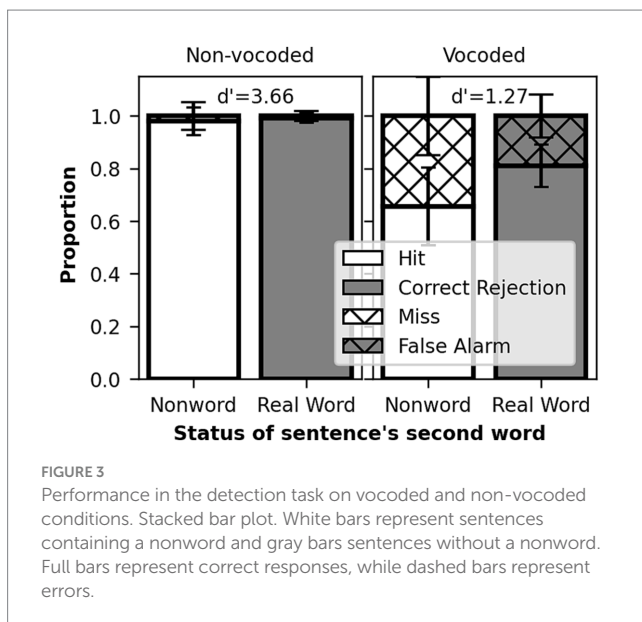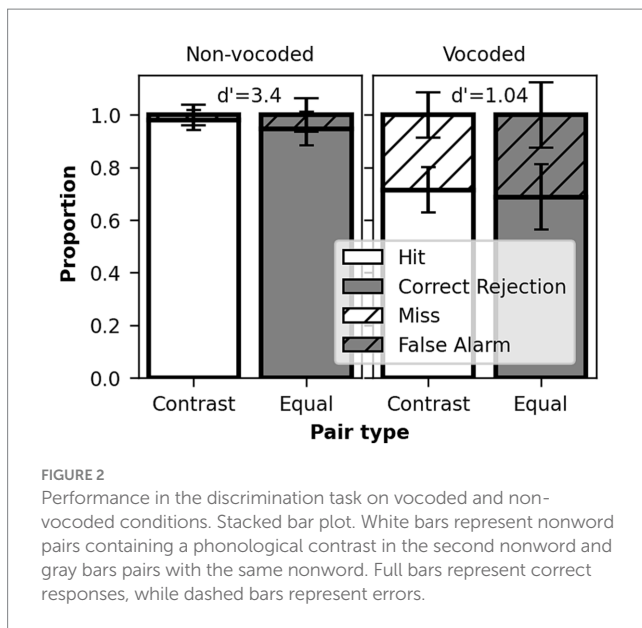
2.5.3. Inferential analysis

Inferential analysis was conducted in Python 3.9, using 'SciPy' (v. 1.7.3) and 'Statsmodels' (v. 0.13.2) packages. Normality in distribution was assessed using Shapiro–Wilk, for the subsequent choice of parametric or nonparametric statistical tests described in the results. Paired comparisons were conducted for mean/median comparison of signal detection performance (d') in vocoded and non-vocoded conditions, and the sequential points in the pupillometric curve in 'yes' versus 'no' tasks. Logistic regression was used to investigate how GCA parameters (intercept, slope, and quadratic term) could be modeled to determine the type of pair ('yes' or 'no') identified. Additionally, effects of vocoding in the pupillometry metrics were analyzed using a linear mixed effect model in a matrix of auditory condition ('vocoded' or 'non-vocoded') and pair type ('yes' or 'no'), with participants attributed as random effects. The inclusion of pair type in the model derives from the assumption that the detection of a phonological contrast in the target word ('yes' tasks) would produce a more prominent response in task evoked pupillometry (Kinzuka et al., 2020).

3. Results

3.1. Performance results

The participants had near ceiling scores on the perception of phonological contrasts for non-vocoded speech, with mean d' scores of 3.39 (SD = 0.56) for the phonological discrimination (Figure 2) and 3.66 (SD = 0.46) for the detection task (Figure 3). For vocoded speech,

FIGURE 2
Performance in the discrimination task on vocoded and non-vocoded conditions. Stacked bar plot. White bars represent nonword pairs containing a phonological contrast in the second nonword and gray bars pairs with the same nonword. Full bars represent correct responses, while dashed bars represent errors.

FIGURE 3
Performance in the detection task on vocoded and non-vocoded conditions. Stacked bar plot. White bars represent sentences containing a nonword and gray bars sentences without a nonword. Full bars represent correct responses, while dashed bars represent errors.

the performance decreased significantly in both tests, with mean d' scores of 1.04 (SD = 0.48) for the phonological discrimination and 1.26 (SD = 0.54) for the detection task. The difference between non-vocoded and vocoded conditions was confirmed by paired comparison t-tests, t (21) = 16.15 for phonological discrimination and t (21) = 15.39 for detection task, $p < 0.001$ for both tasks. Despite lower scores, mean performance was above the 50% chance level in the vocoded condition (mean 70% correct responses for phonological discrimination and 73% correct responses for detection task), confirming that the participants were able to perform both tasks with the vocoded stimuli. Vocoded real-word recognition in the Dantale test had an average accuracy of 34% (SD = 16%). In a simple regression model, the word recognition of vocoded speech alone accounted for over 20% of the variance in the phonological discrimination d' scores, $R^2 = 0.21$, F (1,19) = 4.97, $p = 0.04$, but did not explain the variance in the detection task, $R^2 = 0.01$, F (1,19) = 0.14, $p = 0.71$.

Analyzing the effect of lexical context in the detection of a phonological contrast, Wilcoxon signed-ranks test showed no difference in performance with or without lexical context, when comparing the d' scores of the phonological discrimination and detection tasks, $z = 80.0$, $p = 0.13$. Nevertheless, in the response bias analysis, criterion was located positively at a mean of 0.24 (SD = 0.26) for the detection task, suggesting that participants were biased toward not detecting the nonword despite its presence, while for the phonological discrimination task, criterion was placed much closer to zero, at a mean of −0.04 (SD = 0.18), suggesting no bias on the response.

3.2. Pupillometry responses

The analysis of the pupil data was restricted to trials with correct responses to determine whether successful responses could be differentiated based on pupil dynamics. The aggregated pupillometry response traces across time for both tasks, encompassing data from all participants, are presented in Figure 4 and Figure 5 with respective detailed information in Table 1 and Table 2. In the phonological discrimination task, PPL occurred at a mean of 678 ms (SD = 867) after the presentation of the second word. In the detection task, the PPL for all conditions occurred at a mean of 2.04 s (SD = 1.07) after the onset of the nonword, or after 2.18 s (SD = 1.13) of the onset of the second word for all-real-word sentences when the same alignment was used, which aligns roughly with the offset of the sentence.

3.2.1. Differences in pupil responses due to phonological contrast

In the non-vocoded condition, pupil parameters were sensitive to the presence or absence of the phonological contrast. In the phonological discrimination task, the PPL for pairs without contrast occurred, on average, 217 ms later than pairs with contrast, while PPD values had no significant difference (Table 1). A logistic regression analysis showed no significant effects on the pupil curve intercept, slope, quadratic term, or their interactions, between pairs with and without contrast, $\chi2$ (7, $n = 44$) = 42.6, $p = 0.67$. In the detection task, participants exhibited greater PPD for sentences containing a nonword compared to sentences containing all real words, with no significant difference in PPL (Table 2). The differences in pupil dilation were significant in the interval of 610 ms to 1750 ms after the nonword onset (Figure 5), although the logistic regression analysis did not show any significant differences in the parameters of the fitted curve between trials with and without the phonological contrast, $\chi2$ (7, $n = 37$) = 33.9, $p = 0.45$. No differences between trials with or without the phonological contrast were found for the vocoded stimuli.

3.2.2. Differences in pupil responses due to speech degradation

In contrast to the perceptual results, there was no effect of speech degradation on pupil measures in the phonological discrimination task. This was found when analyzed with a linear mixed effects model that included auditory condition (vocoded or non-vocoded) and pair type (phonological contrast present or absent) as fixed effects, and participant as a random effect (models' marginal R^2s = 0.001 and 0.021, conditional R^2s = 0.394 and 0.311, for PPD and PPL, respectively).

For the detection task, a higher PPD ($\beta = 0.014$, $p = 0.007$) was seen as an effect of speech degradation, when considered in a similar
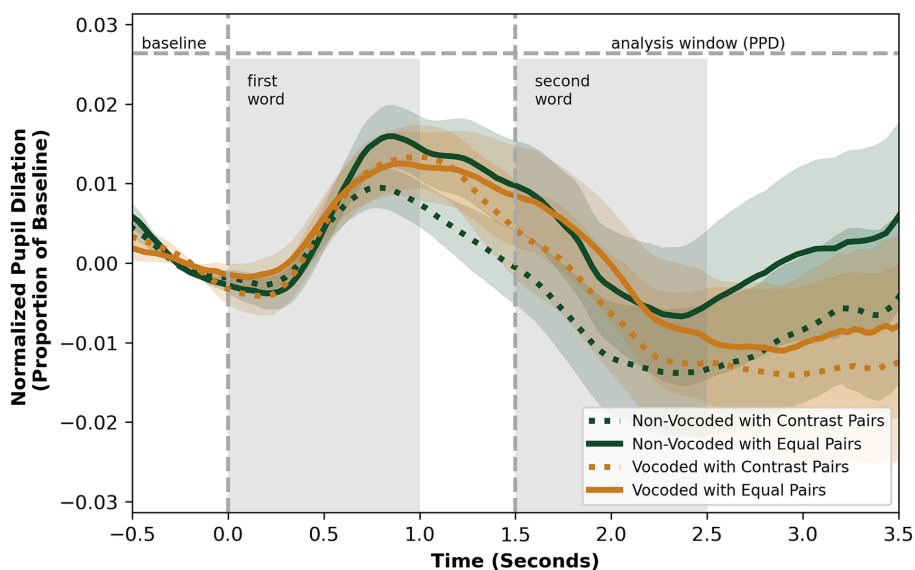
FIGURE 4
Pupil size over time in vocoded and non-vocoded conditions, for the phonological discrimination task, aggregated between participants.
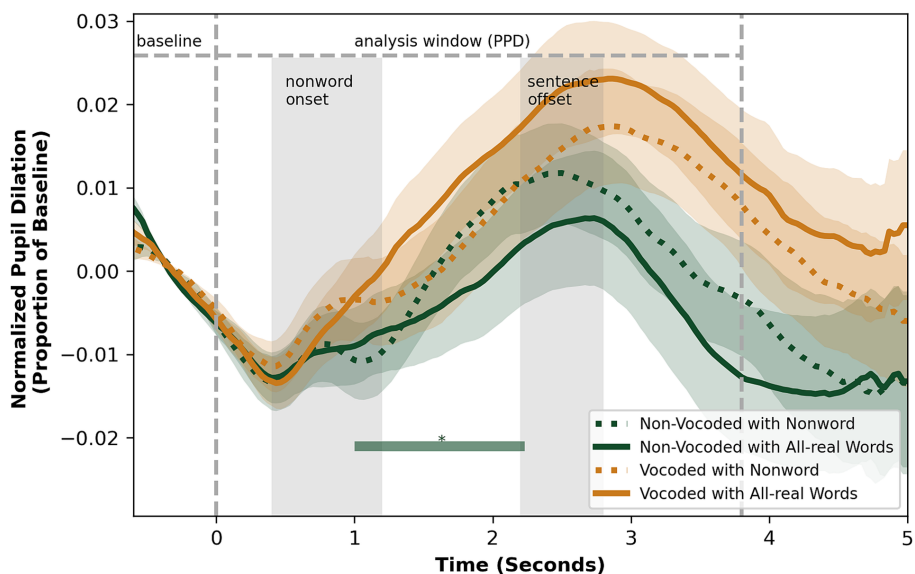

FIGURE 5
Pupil size over time in vocoded and non-vocoded conditions, for the detection task, aggregated between participants. * Timeframes with significant difference between nonword and all-real word sentence types in non-vocoded condition ($p < 0.05$).

condition x pair type linear mixed effects model, with participants as random effects (marginal $R^2 = 0.037$, conditional $R^2 = 0.611$). There was no effect of vocoding on PPL ($\beta = 0.359$, $p = 0.102$, marginal $R^2 = 0.032$, conditional $R^2 = 0.103$).

4. Discussion

This study aimed to investigate the sensitivity of pupil temporal dynamics to multiple levels of auditory processing of phonological information. At the low-level, by analyzing the possibility of detecting phonological discrimination in the absence of lexical context, and at the high-level by directing attention to the phonological contrasts in the presence of context and lack of complete acoustic information.

4.1. Performance results

Performance data showed similar performance in the perception of phonological contrasts in both isolated nonword-pairs and nonwords embedded in sentences. Performance was equally affected by speech degradation, as indicated by changes in accuracy and *d'*

TABLE 1 Pupillometry measures for the phonological discrimination task in the non-vocoded and vocoded conditions, for each pair type.

Condition		Contrast		Equal		t (18)/z	p	n
		M	SD	M	SD			
Non-vocoded	PPD (%)	0.13	0.34	0.24	0.41	-1.2[a]	0.247	19
	PPL (ms)	516	811	787	926	12.5[b]	**0.037***	19
Vocoded	PPD (%)	0.22	0.46	0.17	0.34	1.06[a]	0.303	19
	PPL (ms)	592	844	817	919	14.5[b]	0.054	19

Statistical tests: [a] paired t test; [b] Wilcoxon signed ranks test; PPD = peak pupil dilation; PPL = peak pupil latency.
*p < 0.05.

TABLE 2 Pupillometry measures for the detection task in the non-vocoded and vocoded conditions, for each pair type.

Condition		Nonword		All-real		t (21)/z	p	n
		M	SD	M	SD			
Non-vocoded	PPD (%)	0.20	0.36	0.05	0.37	2.32[a]	**0.030***	22
	PPL (ms)	1844	1,090	2013	1,190	96.0[b]	0.498	22
Vocoded	PPD (%)	0.25	0.39	0.28	0.42	−122.0[b]	0.898	22
	PPL (ms)	2,233	1,030	2,342	1,067	−0.31[a]	0.754	22

Statistical tests: [a] paired t test; [b] Wilcoxon signed ranks test; PPD = peak pupil dilation; PPL = peak pupil latency.
*p < 0.05.

scores in the vocoded speech condition. Although performance was poorer when the stimuli were degraded with a vocoder, participants were still able to perform the task above-chance, with over 70% accuracy. These results demonstrate participants' ability to utilize both low- and high-level strategies to perform speech tasks effectively. However, it indicates that accuracy and sensitivity alone do not provide sufficient information to differentiate between the type of strategy used by individual participants.

The false alarm rate in the vocoded phonological discrimination task (Figure 2) suggests that perceiving two speakers as producing the same nonword was challenging in the degraded speech condition, although the material used in this study has been evaluated as not containing ambiguous phonemes in unprocessed speech (Nielsen and Dau, 2019). False alarms occur when pairs were mistakenly perceived as having a contrast when they did not contain a contrast (i.e., two different speakers producing the same word), revealing a failure to perceive stable signal characteristics during phoneme recognition. Moreover, single-word recognition under vocoded condition appeared as a predictive factor for phonological discrimination, indicating similarity in the tasks' underlying processes. Participants in both tasks were forced to rely on the variable acoustic characteristics of the speech signal to make phonological decisions, and failures occurred unbiased, regardless of the presence or absence of contrast.

Previous studies of phoneme confusion, employing similar 8-channel noise vocoders, have documented higher consonant recognition performance compared to the results observed in this study (Friesen et al., 2001; Xu et al., 2005; Zhou et al., 2010; Jahn et al., 2019; Goupell et al., 2020). These prior studies reported consonant recognition accuracy ranging from approximately 60% (Zhou et al., 2010) in 'consonant-vowel' contexts, to 68% for monosyllables (Goupell et al., 2020) and around 80% (Xu et al., 2005; Jahn et al., 2019) within 'vowel-consonant-vowel' contexts. However, those studies have used a closed set of syllables or words for the consonant recognition task. The open-set word recognition used in our study was more difficult for participants when attempting to identify the target

word. The open-set task increased the number of potential responses, which enhances the activation of neighboring words in the word recognition task. Moreover, for the discrimination of contrasts, the vocoder might be more detrimental to the identification of initial consonants rather than medial consonants, as the highest accuracies were reported in studies using medial consonant identification (Friesen et al., 2001; Jahn et al., 2019). In medial positions, the transition information from vowel to consonant is readily available and contributes to phoneme recognition (Xu et al., 2005). Therefore, the participants' ability to predict the consonants may have been compromised in our study, shown by the reduced accuracy scores in the phonological discrimination task.

As expected, the presence of context led to a bias toward reporting nonwords as real words in the vocoded detection task, causing the participants to ignore or miss the phonological contrast. A degraded signal amplifies the perceptual bias in phonological perception, increasing the reliance on non-acoustic information such as lexical information and context when categorizing phonological contrasts (Gianakas and Winn, 2019; Vickery et al., 2022; Winn and Teece, 2022), producing the effects observed.

4.2. Pupillometry responses

4.2.1. Differences in pupil responses due to phonological contrast

The presence of a phonological contrast did not elicit higher pupil dilation in the phonological discrimination task, as it would be expected in a presence of a variant stimuli (Wagner et al., 2016; Kinzuka et al., 2020). The differences here can be attributed to the demands of the tasks. The simplicity of the forced-choice task might not have elicited sufficient differences in the demand for cognitive processing to capture the effect of the phonological contrast. Additionally, in contrast to previous studies which used words as material for discrimination, in our study the participants could not

use lexical information to support the decision regarding the change in the phoneme category. Therefore, their judgment was forced to occur solely at the phonological level. The absence of a significant difference in pupil parameters suggests that pupil dynamics may be more sensitive to higher-level cognitive and language processing, as to lexical categorization (Kamp and Donchin, 2015), rather than lower-level phonological categorization.

Moreover, the pupil response to phonological contrasts may be indistinguishable from the response for the perception of acoustic contrasts. The contrast between two speakers in our paradigm, one in each token in the nonword-pair, was done to ensure that discrimination was occurring at a phonological rather than acoustical level. It is known that different speakers possess a natural variability in multiple acoustic domains as voice-onset-time, vowel formants, consonant intensity, among others (Allen et al., 2003; Christiansen and Henrichsen, 2011). Therefore, identifying two nonwords as the same would require their processing at the phonological level. However, as the pupil dilates for acoustic deviants, such as pure tones and noise varying in frequency (Liao et al., 2016; Selezneva et al., 2021), pupil dilation could also be an index of the processing of the dynamic acoustic characteristics of speech in an effort to solve ambiguity caused by interspeaker variations in phoneme production and boundaries (Lewis and Bidelman, 2020; Winn, 2020; Reese and Reinisch, 2022; Yu, 2022). Such a response to acoustic differences would explain the comparable PPDs recorded for both pairs with and without phonological contrast, since for both types of pairs the acoustic variability was present.

Interestingly, in the phonological discrimination task, PPLs were shorter for pairs with phonological contrast than for pairs without contrast. Koelewijn et al. (2017) describe the PPL as a measure of the speed of cognitive processing, with shorter latencies indicating faster cognitive processing or the need for processing less information. One explanation for our results is that to correctly identify a phonological contrast, the participant would only need to identify the first phoneme of the second word in the pair, but to correctly identify the absence of a contrast required the processing of the whole nonword in a pair. Therefore, a decision could be taken quicker with far less information for pairs with contrast.

The presence of context, in the detection task, led to higher PPD in sentences containing a phonological altered word (nonword). This effect was expected as it had been previously reported by Wagner et al. (2016) and Winn and Teece (2021, 2022). These studies found that substituted and distorted phonemes within words in a sentence lead to steeper pupil dilation. As discussed in Winn and Teece (2022), the presence of sentence context makes it difficult to determine if the higher dilation occurs due to increased cognitive demand for sentence processing introduced by the ambiguous lexical entry, or due to the detection of the phonological contrast. However, the absence of difference in the results of the phonological discrimination task suggests that the pupil response may be more closely linked to the violation of the lexical expectation rather than the phonological contrast.

Remarkably, in our study, participants were not asked to process the whole sentence in any manner (they did not repeat it back, nor derived its meaning). Therefore, it could be expected that after detecting the nonword in the second position of the sentence, the participants' demand for processing would immediately decrease, which should have resulted in a reduction in the pupil size. Yet, the observed pupil behavior indicates that the whole sentence was processed before the response was given. Despite the different protocols used, these results are consistent with Winn and Teece (2021, 2022), in which participants were asked to repeat the whole sentence back to the experimenter. These findings suggest that, despite being instructed to track individual words in the sentence, listeners may have used the whole sentence context to make decisions regarding the presence or absence of the phonological contrast. As an anecdotal report, during the experiment session, several participants reported attempting to 'repair' the nonword or 'figure out the correct word'.

4.2.2. Differences in pupil responses due to speech degradation

The results in the vocoded condition support the argument that the pupil response reflects processing at the lexical and sentence level. The high accuracy scores for identifying the presence of a nonword within a sentence shows that participants were able to detect the phonological alteration despite the vocoded speech, indicating that phonological discrimination was occurring at a low-level. However, the lack of difference in the pupil parameters between sentences with and without phonologically modified words suggests that the pupil response captured the increase in cognitive processing required to understand the vocoded sentences, rather than the detection of a phonological contrast. Furthermore, the trend of interpreting nonwords as real words in the performance results suggests that the participants were likely attempting phonological restoration throughout the vocoded experiment. In other words, it is possible that the absence of differences in the pupillary response between stimuli with and without phonological contrast reflects the registration of a different type of response besides the detection of the contrast. The physiological mechanisms underlying pupil dilation are also involved in the process of decision-making (Kafkas and Montaldi, 2018). As such, when decisions require greater cognitive processing and memory demand, pupil size increases. It is important to note that the signal restoration of the vocoded stimuli comes at a cost even for real words (Winn et al., 2015; Balling et al., 2017). This global response, which is related to the processing of the auditory stimulus as a whole, may be more pronounced than the response to the detection of the phonological contrast, thereby masking its signal.

Another possible explanation for the lack of difference in pupil metrics between stimuli with or without phonological contrast is that errors in detecting the contrast may have occurred at different moments in the stimuli presentation. Since participants were not instructed about the possible location of the phonological contrast, it was not possible to track the exact moment when errors occurred. As a result, the effect of the phonological contrast may have been distributed across the time series average (Winn and Teece, 2021), which could not be tracked by our analysis.

4.3. Study limitations

As in any forced-choice task, the methodology used opens the possibility for participants to 'guess' the responses. This effect can be considered during the signal detection analysis of the performance but might influence the amplitude and morphology of the pupil responses. Responses based on chance, with low or no processing of the stimulus, can contaminate the time-series average during pupil

analysis and effects be missed. Additionally, pupil responses are modulated by the sympathetic nervous system, which can be influenced by a range of factors such as engagement, fatigue, or self-perception of performance (Hopstaken et al., 2015; Zekveld et al., 2018; McGarrigle et al., 2021). Although we attempted to counterbalance for fatigue effects by randomizing the order of the presentation of the tasks and stimuli, it is possible that the low scores in speech perception, achieved in the vocoded speech condition, have led to disengagement from the task, which would be reflected in an overall reduction of pupil dilation (Hopstaken et al., 2015; Ohlenforst et al., 2017).

It is worth noting that the characteristics of the phonological contrast in the phonological discrimination task and the detection task were not the same. While in the phonological discrimination task the contrast was defined by a change in one production feature, multiple production features were modified in the detection task. In terms of acoustic differences, this might mean that the acoustic degradation would affect different aspects of the phonological perception in each task (Xu et al., 2005; Zhou et al., 2010). Furthermore, it raises the possibility that pupil dilation would be sensitive to the distance between the expected stimulus and the contrast, as previously observed for non-speech stimuli (Liao et al., 2016; Winn and Teece, 2021).

Furthermore, participants in our study were exposed to only a brief practice session with the vocoded stimuli. While this training was conducted similarly as previous studies (Hervais-Adelman et al., 2011; Winn et al., 2015), adapting to vocoded speech may require longer practice (Hervais-Adelman et al., 2011). Thus, it is possible that the immediate results produced by the spectral degradation would not have been sustained in a longer task, which would have induced phonological accommodation and potentially have led to better speech recognition (Jesse, 2021).

5. Conclusion

The present study offers insights on the pupil temporal dynamics from the processing of phonological information. The lack of differences in the pupil dilation to the presence of a phonological contrast in lexically decontextualized nonwords (phonological discrimination task) could suggest that pupil dynamics are more sensitive to higher-level cognitive and language processing, such as lexical categorization, rather than lower-level phonological categorization. Nevertheless, the pupil response to phonological contrasts may overlap with responses to acoustic differences, indicating that pupil dilation may reflect the processing of dynamic acoustic characteristics of speech.

In the presence of lexical/contextual information (detection task), phonological contrasts led to higher pupil dilation. This increase in pupil dilation could be attributed either to an increase in cognitive demand for processing a sentence containing a nonword, or to a response to the detection of the phonological contrast. The inability to distinguish between high and low-level processing in the detection task stemmed from participants' apparent reliance on sentence context when making decisions about phonological contrasts, despite explicit instructions to track individual words.

These findings bring important considerations to the use of pupillometry when investigating phonological perception in the presence of lexical meaning or acoustic variability. Further research is needed to gain a comprehensive understanding of the intricate interactions among acoustic, phonological, and linguistic factors and their influence on pupil dynamics during speech perception.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The requirement of ethical approval was waived by the Scientific Ethics Committees, Center for Regional Development, Capital Region - Denmark for the studies involving humans because the Scientific Ethics Committees, Center for Regional Development, Capital Region - Denmark considered it to be a study in the social domain. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

Funding

Acknowledgments

Conflict of interest

JC, FP, EN, and KF were employed by the company Oticon A/S, Smørum, Denmark, while this study was conducted.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abada, S. H., Baum, S. R., and Titone, D. (2008). The effects of contextual strength on phonetic identification in younger and older listeners. *Exp. Aging Res.* 34, 232–250. doi: 10.1080/03610730802070183

Allen, J. S., Miller, J. L., and DeSteno, D. (2003). Individual talker differences in voice-onset-time. *J. Acoust. Soc. Am.* 113, 544–552. doi: 10.1121/1.1528172

Bala, A. D. S., Whitchurch, E. A., and Takahashi, T. T. (2020). Human auditory detection and discrimination measured with the pupil dilation response. *J. Assoc. Res. Otolaryngol.* 21, 43–59. doi: 10.1007/s10162-019-00739-x

Balling, L. W., Morris, D. J., and Tøndering, J. (2017). Investigating lexical competition and the cost of phonemic restoration. *J. Acoust. Soc. Am.* 142, 3603–3612. doi: 10.1121/1.5017603

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., et al. (2008a). Early vocabulary development in Danish and other languages: a CDI-based comparison. *J. Child Lang.* 35, 619–650. doi: 10.1017/S0305000908008714

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., et al. (2008b). The Danish communicative developmental inventories: validity and main developmental trends. *J. Child Lang.* 35, 651–669. doi: 10.1017/S0305000907008574

Boersma, P., and Weenink, D. (1992). *Praat: doing phonetics by computer*. Available at: https://www.praat.org (Accessed January 2, 2022).

Borsky, S., Tuller, B., and Shapiro, L. P. (1998). "How to milk a coat:" the effects of semantic and acoustic information on phoneme categorization. *J. Acoust. Soc. Am.* 103, 2670–2676. doi: 10.1121/1.422787

Burg, E. A., Thakkar, T., Fields, T., Misurelli, S. M., Kuchinsky, S. E., Roche, J., et al. (2021). Systematic comparison of trial exclusion criteria for Pupillometry data analysis in individuals with single-sided deafness and Normal hearing. *Trends Hear.* 25:23312165211013256. doi: 10.1177/23312165211013256

Casserly, E. D., and Pisoni, D. B. (2010). Speech perception and production. *Wiley Interdisc. Rev. Cogn. Sci.* 1, 629–647. doi: 10.1002/wcs.63

Christiansen, T. U., and Henrichsen, P. J. (2011). *Objective evaluation of consonant-vowel pairs produced by native speakers of Danish*. European Acoustics Association, EAA. Madrid

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

Coleman, J. (2003). Discovering the acoustic correlates of phonological contrasts. *J. Phon.* 31, 351–372. doi: 10.1016/j.wocn.2003.10.001

Dan, E. L., Dînşoreanu, M., and Mureşan, R. C. (2020). Accuracy of six interpolation methods applied on pupil diameter data. In 2020 IEEE international conference on automation, quality and testing, robotics (AQTR), 1–5. IEEE. Cluj-Napoca, Romania

Dorman, M. F., Loizou, P. C., and Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *J. Acoust. Soc. Am.* 102, 2403–2411. doi: 10.1121/1.419603

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., and Bunge, S. A. (2017). Beyond eye gaze: what else can eyetracking reveal about cognition and cognitive development? *Dev. Cogn. Neurosci.* 25, 69–91. doi: 10.1016/j.dcn.2016.11.001

Elberling, C., Ludvigsen, C., and Lyregaard, P. E. (1989). Dantale: a new Danish speech material. *Scand. Audiol.* 18, 169–175. doi: 10.3109/01050398909070742

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J. Acoust. Soc. Am.* 110, 1150–1163. doi: 10.1121/1.1381538

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *J. Exp. Psychol. Hum. Percept. Perform.* 6, 110–125. doi: 10.1037//0096-1523.6.1.110

Gianakas, S. P., and Winn, M. B. (2019). Lexical bias in word recognition by cochlear implant listeners. *J. Acoust. Soc. Am.* 146, 3373–3383. doi: 10.1121/1.5132938

Goupell, M. J., Draves, G. T., and Litovsky, R. Y. (2020). Recognition of vocoded words and sentences in quiet and multi-talker babble with children and adults. *PLoS One* 15:e0244632. doi: 10.1371/journal.pone.0244632

Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., and Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 283–295. doi: 10.1037/a0020772

Hopstaken, J. F., van der Linden, D., Bakker, A. B., and Kompier, M. A. J. (2015). The window of my eyes: task disengagement and mental fatigue covary with pupil dynamics. *Biol. Psychol.* 110, 100–106. doi: 10.1016/j.biopsycho.2015.06.013

Iverson, P. (2003). Evaluating the function of phonetic perceptual phenomena within speech recognition: an examination of the perception of /d/–/t/ by adult cochlear implant users. *J. Acoust. Soc. Am.* 113, 1056–1064. doi: 10.1121/1.1531985

Jahn, K. N., DiNino, M., and Arenberg, J. G. (2019). Reducing simulated channel interaction reveals differences in phoneme identification between children and adults with Normal hearing. *Ear Hear.* 40, 295–311. doi: 10.1097/AUD.0000000000000615

Jesse, A. (2021). Sentence context guides phonetic retuning to speaker idiosyncrasies. *J. Exp. Psychol. Learn. Mem. Cogn.* 47, 184–194. doi: 10.1037/xlm0000805

Johnsrude, I. S., and Rodd, J. M. (2016). "Chapter 40 - factors that increase processing demands when listening to speech" in *Neurobiology of language*. eds. G. Hickok and S. L. Small (San Diego: Academic Press), 491–502.

Kadem, M., Herrmann, B., Rodd, J. M., and Johnsrude, I. S. (2020). Pupil dilation is sensitive to semantic ambiguity and acoustic degradation. *Trends Hear.* 24:2331216520964068. doi: 10.1177/2331216520964068

Kafkas, A., and Montaldi, D. (2018). How do memory systems detect and respond to novelty? *Neurosci. Lett.* 680, 60–68. doi: 10.1016/j.neulet.2018.01.053

Kamp, S.-M., and Donchin, E. (2015). ERP and pupil responses to deviance in an oddball paradigm. *Psychophysiology* 52, 460–471. doi: 10.1111/psyp.12378

Kinzuka, Y., Minami, T., and Nakauchi, S. (2020). Pupil dilation reflects English /l//r/ discrimination ability for Japanese learners of English: a pilot study. *Sci. Rep.* 10:8052. doi: 10.1038/s41598-020-65020-1

Koelewijn, T., Versfeld, N. J., and Kramer, S. E. (2017). Effects of attention on the speech reception threshold and pupil response of people with impaired and normal hearing. *Hear. Res.* 354, 56–63. doi: 10.1016/j.heares.2017.08.006

Kramer, S. E., Lorens, A., Coninx, F., Zekveld, A. A., Piotrowska, A., and Skarzynski, H. (2013). Processing load during listening: the influence of task characteristics on the pupil response. *Lang. Cogn. Process.* 28, 426–442. doi: 10.1080/01690965.2011.642267

Kret, M. E., and Sjak-Shie, E. E. (2019). Preprocessing pupil size data: guidelines and code. *Behav. Res. Methods* 51, 1336–1342. doi: 10.3758/s13428-018-1075-y

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 979–1000. doi: 10.1098/rstb.2007.2154

Kutas, M., and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163. doi: 10.1038/307161a0

Lewis, G. A., and Bidelman, G. M. (2020). Autonomic nervous system correlates of speech categorization revealed through Pupillometry. *Front. Neurosci.* 13, 1–10. doi: 10.3389/fnins.2019.01418

Liao, H.-I., Yoneya, M., Kidani, S., Kashino, M., and Furukawa, S. (2016). Human pupillary dilation response to deviant auditory stimuli: effects of stimulus properties and voluntary attention. *Front. Neurosci.* 10:43. doi: 10.3389/fnins.2016.00043

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279

Macmillan, N. A., and Creelman, C. D. (2005). *Detection theory: a user's guide*, 2nd ed. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Mathôt, S., Schreij, D., and Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav. Res. Methods* 44, 314–324. doi: 10.3758/s13428-011-0168-7

Mathôt, S., and Vilotijević, A. (2022). Methods in cognitive pupillometry: design, preprocessing, and statistical analysis. *Behav. Res. Methods* 55, 3055–3077. doi: 10.3758/s13428-022-01957-7

McGarrigle, R., Rakusen, L., and Mattys, S. (2021). Effortful listening under the microscope: examining relations between pupillometric and subjective markers of effort and tiredness from listening. *Psychophysiology* 58:e13703. doi: 10.1111/psyp.13703

Micula, A., Rönnberg, J., Książek, P., Murmu Nielsen, R., Wendt, D., Fiedler, L., et al. (2022). A glimpse of memory through the eyes: pupillary responses measured during encoding reflect the likelihood of subsequent memory recall in an auditory free recall test. *Trends Hear.* 26:233121652211305. doi: 10.1177/23312165221130581

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Nielsen, J. B., and Dau, T. (2019). A Danish nonsense word corpus for phoneme recognition measurements. *Acta. Acust. United Acust.* 105, 183–194. doi: 10.3813/AAA.919299

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., and Lunner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hear. Res.* 365, 90–99. doi: 10.1016/j.heares.2018.05.003

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., et al. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hear. Res.* 351, 68–79. doi: 10.1016/j.heares.2017.05.012

Phatak, S. A., and Grant, K. W. (2014). Phoneme recognition in vocoded maskers by normal-hearing and aided hearing-impaired listeners. *J. Acoust. Soc. Am.* 136, 859–866. doi: 10.1121/1.4889863

Pittman, A. L., and Schuett, B. C. (2013). Effects of Semantic and Acoustic Context on Nonword Detection in Children With Hearing Loss. Ear & Hearing. 34, 213–220. doi: 10.1097/AUD.0b013e31826e5006

Pittman, A. L., Stewart, E. C., Odgear, I. S., and Willman, A. P. (2017). Detecting and learning new words: the impact of advancing age and hearing loss. *Am. J. Audiol.* 26, 318–327. doi: 10.1044/2017_AJA-17-0025

Reese, H., and Reinisch, E. (2022). Cognitive load does not increase reliance on speaker information in phonetic categorization. *JASA Express Lett.* 2:055203. doi: 10.1121/10.0009895

Repp, B. H. (1982). Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychol. Bull.* 92, 81–110. doi: 10.1037/0033-2909.92.1.81

Romberg, A. R., and Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 906–914. doi: 10.1002/wcs.78

Selezneva, E., Brosch, M., Rathi, S., Vighneshvel, T., and Wetzel, N. (2021). Comparison of pupil dilation responses to unexpected sounds in monkeys and humans. *Front. Psychol.* 12:754604. doi: 10.3389/fpsyg.2021.754604

Seropian, L., Ferschneider, M., Cholvy, F., Micheyl, C., Bidet-Caulet, A., and Moulin, A. (2022). Comparing methods of analysis in pupillometry: application to the assessment of listening effort in hearing-impaired patients. *Heliyon* 8:e09631. doi: 10.1016/j.heliyon.2022.e09631

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instrum. Comput.* 31, 137–149. doi: 10.3758/BF03207704

Steinhauer, K., and Connolly, J. F. (2008). "Event-related potentials in the study of language" in *Handbook of the Neuroscience of Language*. eds. B. Stemmer and H. Whitaker (Canada: Elsevier), 91–104.

Stenfelt, S., and Rönnberg, J. (2009). The signal-cognition interface: interactions between degraded auditory signals and cognitive processes. *Scand. J. Psychol.* 50, 385–393. doi: 10.1111/j.1467-9450.2009.00748.x

Sulas, E., Hasan, P.-Y., Zhang, Y., and Patou, F. (2022). *Streamlining experiment design in cognitive hearing science using OpenSesame. Behav. Res. Methods.*. 55, 1965–1979, doi: 10.3758/s13428-022-01886-5

Trau-Margalit, A., Fostick, L., Harel-Arbeli, T., Nissanholtz Gannot, R., and Taitelbaum-Swead, R. (2023). Speech recognition in noise task among children and young-adults: a pupillometry study. *Front. Psychol.* 14:1188485. doi: 10.3389/fpsyg.2023.1188485

Vickery, B., Fogerty, D., and Dubno, J. R. (2022). Phonological and semantic similarity of misperceived words in babble: effects of sentence context, age, and hearing loss. *J. Acoust. Soc. Am.* 151, 650–662. doi: 10.1121/10.0009367

Virtala, P., Partanen, E., Tervaniemi, M., and Kujala, T. (2018). Neural discrimination of speech sound changes in a variable context occurs irrespective of attention and explicit awareness. *Biol. Psychol.* 132, 217–227. doi: 10.1016/j.biopsycho.2018.01.002

Wagner, A. E., Toffanin, P., and Başkent, D. (2016). The timing and effort of lexical access in natural and degraded speech. *Front. Psychol.* 7:398. doi: 10.3389/fpsyg.2016.00398

Wendt, D., Dau, T., and Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Front. Psychol.* 7:345. doi: 10.3389/fpsyg.2016.00345

Winn, M. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and Cochlear implants. *Trends Hear.* 20:2331216516669723. doi: 10.1177/2331216516669723

Winn, M. B. (2020). Accommodation of gender-related phonetic differences by listeners with cochlear implants and in a variety of vocoder simulations. *J. Acoust. Soc. Am.* 147, 174–190. doi: 10.1121/10.0000566

Winn, M. B. (2021). *Vocoder: vocode all selected sounds in the objects list or all sounds in a specified folder*. Available at: http://www.mattwinn.com/praat/vocode_all_selected_v45.txt (Accessed August 2, 2022).

Winn, M. B., Chatterjee, M., and Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: effects of spectral degradation and electric hearing. *J. Acoust. Soc. Am.* 131, 1465–1479. doi: 10.1121/1.3672705

Winn, M. B., Edwards, J. R., and Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear.* 36, e153–e165. doi: 10.1097/AUD.0000000000000145

Winn, M. B., and Teece, K. H. (2021). Listening effort is not the same as speech intelligibility score. *Trends Hear.* 25:23312165211027688. doi: 10.1177/23312165211027688

Winn, M. B., and Teece, K. H. (2022). Effortful listening despite correct responses: the cost of mental repair in sentence recognition by listeners with Cochlear implants. *J. Speech Lang. Hear. Res.* 65, 3966–3980. doi: 10.1044/2022_JSLHR-21-00631

Winn, M. B., Wendt, D., Koelewijn, T., and Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: an introduction for those who want to get started. *Trends Hear.* 22:2331216518800869. doi: 10.1177/2331216518800869

Xu, L., Thompson, C. S., and Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *J. Acoust. Soc. Am.* 117, 3255–3267. doi: 10.1121/1.1886405

Yu, A. C. L. (2022). Perceptual cue weighting is influenced by the listener's gender and subjective evaluations of the speaker: the case of English stop voicing. *Front. Psychol.* 13:840291. doi: 10.3389/fpsyg.2022.840291

Zekveld, A. A., Koelewijn, T., and Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: current state of knowledge. *Trends Hear.* 22:2331216518777174. doi: 10.1177/2331216518777174

Zhang, Y., Malaval, F., Lehmann, A., and Deroche, M. L. D. (2022). Luminance effects on pupil dilation in speech-in-noise recognition. *PLoS One* 17:e0278506. doi: 10.1371/journal.pone.0278506

Zhou, N., Xu, L., and Lee, C.-Y. (2010). The effects of frequency-place shift on consonant confusion in cochlear implant simulations. *J. Acoust. Soc. Am.* 128, 401–409. doi: 10.1121/1.3436558