frontiers in Neuroscience

BRIEF RESEARCH REPORT
published: 10 January 2020
doi: 10.3389/fnins.2019.01418

Autonomic Nervous System Correlates of Speech Categorization Revealed Through Pupillometry

Gwyneth A. Lewis[1,2] *and Gavin M. Bidelman*[1,2,3]*

[1] *Institute for Intelligent Systems, The University of Memphis, Memphis, TN, United States,* [2] *School of Communication Sciences and Disorders, The University of Memphis, Memphis, TN, United States,* [3] *Department of Anatomy and Neurobiology, University of Tennessee Health Sciences Center, Memphis, TN, United States*

Human perception requires the many-to-one mapping between continuous sensory elements and discrete categorical representations. This grouping operation underlies the phenomenon of categorical perception (CP)—the experience of perceiving discrete categories rather than gradual variations in signal input. Speech perception requires CP because acoustic cues do not share constant relations with perceptual-phonetic representations. Beyond facilitating perception of unmasked speech, we reasoned CP might also aid the extraction of target speech percepts from interfering sound sources (i.e., noise) by generating additional perceptual constancy and reducing listening effort. Specifically, we investigated how noise interference impacts cognitive load and perceptual identification of unambiguous (i.e., categorical) vs. ambiguous stimuli. Listeners classified a speech vowel continuum (/u/-/a/) at various signal-to-noise ratios (SNRs [unmasked, 0 and −5 dB]). Continuous recordings of pupil dilation measured processing effort, with larger, later dilations reflecting increased listening demand. Critical comparisons were between time-locked changes in eye data in response to unambiguous (i.e., continuum endpoints) tokens vs. ambiguous tokens (i.e., continuum midpoint). Unmasked speech elicited faster responses and sharper psychometric functions, which steadily declined in noise. Noise increased pupil dilation across stimulus conditions, but not straightforwardly. Noise-masked speech modulated peak pupil size (i.e., [0 and −5 dB] > unmasked). In contrast, peak dilation latency varied with both token and SNR. Interestingly, categorical tokens elicited earlier pupil dilation relative to ambiguous tokens. Our pupillary data suggest CP reconstructs auditory percepts under challenging listening conditions through interactions between stimulus salience and listeners' internalized effort and/or arousal.

Keywords: pupillometry, categorical perception, speech-in-noise (SIN) perception, listening effort, eye behavior

OPEN ACCESS

Edited by:
Yi Du,
Institute of Psychology (CAS), China

Reviewed by:
Alexander Francis,
Purdue University, United States
Jackson Everett Graves,
University of Minnesota Twin Cities, United States

***Correspondence:**
Gavin M. Bidelman
gmbdlman@memphis.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *04 June 2019*
Accepted: *16 December 2019*
Published: *10 January 2020*

Citation:
Lewis GA and Bidelman GM (2020) Autonomic Nervous System Correlates of Speech Categorization Revealed Through Pupillometry. Front. Neurosci. 13:1418. doi: 10.3389/fnins.2019.01418

INTRODUCTION

Virtually all sensory signals vary along a physical continuum, yet, we tend to perceive them as discrete perceptual objects. Such categorical perception (CP) deciphers meaningful patterns in complex sensory input by organizing information into coherent groups (equivalence classes) (Goldstone and Hendrickson, 2010). Nowhere is this phenomenon more robustly demonstrated

than in speech perception. When listeners hear tokens from a phonetic continuum, their discriminability is very good for sounds straddling the category boundary near the midpoint, but very poor for sounds on the same side (Liberman et al., 1967; Pisoni, 1973; Harnad, 1987; Pisoni and Luce, 1987; Bidelman et al., 2013). CP streamlines speech processing by emphasizing acoustic contrasts between- rather than within-phoneme categories (Myers and Swan, 2012), presumably by weighting cues for comparison against internalized templates of a person's native speech sounds (Kuhl, 1991; Iverson et al., 2003; Guenther et al., 2004; Bidelman and Lee, 2015).

Neuroimaging work has revealed neural processes leading up to categorical decisions (Sharma and Dorman, 1999; Binder et al., 2004; Chang et al., 2010; Zhang et al., 2011; Bidelman et al., 2013; Bidelman and Lee, 2015). In the auditory sciences, research has associated measures of perceptual performance and "listening effort," which is the deliberate allocation of (available) mental resources to overcome goals when carrying out a listening task (for review see, Zekveld et al., 2018). Under the Framework for Understanding Effortful Listening (FUEL), listening effort is determined by the combined effect of input-demands (e.g., signal quality) and internal factors (e.g., arousal, attention, and motivation) (Pichora-Fuller et al., 2016). Accounting for the latter is crucial interpreting apparent task-related differences.

Diverse experimental techniques have shown that noise degradation has robust consequences for perceptual performance (e.g., Gatehouse and Gordon, 1990), short-term memory performance (e.g., Heinrich et al., 2008), neural activity (e.g., Scott et al., 2000), and pupil reactivity (e.g., Zekveld et al., 2011). Acoustic noise burdens cognitive load, but speech intelligibility is not always straightforwardly predicted by signal-to-noise-ratio (SNR) (for review see, Bidelman, 2017). Under the Ease of Language Understanding (ELU) model (Rönnberg et al., 2013), acoustic input that deviates from a listener's long-term phonological memory store requires additional cognitive resources for recognition, including working memory and executive functions. The degree to which listeners engage explicit cognitive processes is thought to reflect task-related listening effort, however, cognitive resources and intrinsic motivation may be insufficient for recognition when the mismatch between percept and expectation is too extreme (Ohlenforst et al., 2017).

Segregating a speech signal from acoustic noise is cognitively demanding, drawing on resources for encoding that are normally used for other processes (Cousins et al., 2014). Mechanisms for signal separation might be more readily engaged when category boundaries are particularly noisy (Livingston et al., 1998). Neuroimaging data indicates that the brain processes competing sound streams within the same neural pathways, but devotes more attention to the target stream (Evans et al., 2016). Our recent electrophysiological study found that neural activity was not only stronger for category (unambiguous) relative to non-category (ambiguous) speech sounds but the former was more invariant to noise interference, suggesting CP promotes robust speech perception by "sharpening" category members in noisy feature space (Bidelman et al., 2019b).

Because underlying processes are difficult to measure behaviorally, researchers have assessed listening effort with indirect measurement techniques. For example, eyetracking offers an objective glimpse into real-time speech processing (Ben-David et al., 2011) not captured by behavioral measures and self-reports (Wendt et al., 2016). One non-volitional indicator of cognitive processes is pupil reactivity (pupillometry) (see Naylor et al., 2018). Studies have reported close relations between fluctuations in pupil diameter and underlying neural mechanisms (for review see, Eckstein et al., 2017). Pupil diameter increases with momentary cognitive demands (Kahneman and Beatty, 1966) and correlates closely with neuronal activity from the locus coeruleus, which is the principal brain site for synthesizing norepinephrine (i.e., arousal) (Aston-Jones and Cohen, 2005). Thus, pupil diameter indirectly indicates processes below the threshold of consciousness, which can be modulated by task demands. On a practical note, pupillometry complements other online measures of speech processing, is relatively simple to administer, and can be simultaneously registered with neurophysiological measures (e.g., for review see, Winn et al., 2018).

From the perspective of listening effort, pupillometry is an ideal avenue for investigating the physiological nature and individual differences in speech categorization. Germane to our interests in speech processing, aspects of the pupil response systematically vary with processing load when interpreting languages (Hyönä et al., 1995), speech intelligibility (Zekveld et al., 2010), divided attention during speech listening (Koelewijn et al., 2014), semantic ambiguity (Vogelzang et al., 2016), visual-auditory semantic incongruency (Renner and Wlodarczak, 2017), and pseudoword complexity (López-Ornat et al., 2018). Relevant to this study, researchers have used pupillometry and eyetracking methods to examine how acoustically degraded speech influences listening effort (e.g., Bidelman et al., 2019a; Winn et al., 2015). Findings have been largely consistent: peak pupil dilation and latency systematically increase with decreasing speech intelligibility, but only to the extent that cognitive resources are not overloaded (see section "Discussion") (Zekveld et al., 2010; Zekveld and Kramer, 2014; Wendt et al., 2016; Ohlenforst et al., 2018). Assessing how pupil responses vary with listening effort could reveal how CP reconstructs auditory percepts under challenging listening conditions. Presumably, speech categorization depends on interactions between stimulus salience (Liao et al., 2016) and listeners' internalized effort and/or arousal (for attentional dependence of CP, see Bidelman and Walker, 2017).

Here, we investigated how noise interference impacts cognitive load during perceptual identification of speech. Members of speech sound continua were presented in varying levels of noise to parametrically manipulate listening effort above and beyond that needed to classify unambiguous and ambiguous speech. Using pupillometry, we acquired continuous recordings of pupil dilation as a proxy of listening effort. If the grouping mechanisms of CP aid figure-ground perception of speech, we hypothesized unambiguous phonemes (categories) should elicit less noise-related changes in pupil responses than ambiguous tokens lacking a clear categorical identity. Our data show that the categorical nature of speech not only reduces cognitive load (listening effort) but also assists speech perception in noise degraded environments.

METHODS

Participants

Fifteen young adults (3 males, 12 females; age: M = 24.3, SD = 1.7 years) from The University of Memphis participated in the experiment. All exhibited normal hearing sensitivity (i.e., <20 dB HL thresholds, 250–8000 Hz). Each participant was strongly right-handed (87.0 ± 18.2 laterality index; Oldfield, 1971) and had obtained a collegiate level of education (17.8 ± 1.9 years). Musical training enhances categorical processing and speech-in-noise listening abilities (Bidelman et al., 2014; Yoo and Bidelman, 2019). Consequently, all participants were required to have < 3 years of music training throughout their lifetime (mean years of training: 1.3 ± 1.8 years). All were paid for their time and gave written informed consent in compliance with a protocol approved by the Institutional Review Board at the University of Memphis.

Speech Stimuli and Behavioral Task

We used a synthetic five-step vowel continuum previously used to investigate the neural correlates of CP (see Figure 1 of Bidelman et al., 2013; Bidelman and Walker, 2017). Each token was separated by equidistant linear steps acoustically based on first formant frequency (F1) yet was designed to be perceived categorically from /u/ to /a/. Although vowel sounds are perceived less categorically than other speech sounds (e.g., stop-consonants; Pisoni, 1973, 1975; Altmann et al., 2014), they do not carry intrinsic features upon which to make category judgments (formant transitions in consonants, for example, allow comparisons within the stimulus itself) (for discussion, see Xu et al., 2006). In contrast, steady-state features like the F1 contrast of our static vowels lack an intrinsic reference so categorical hearing of these stimuli necessarily requires acoustic features be matched to the best exemplar in long-term memory (Pisoni, 1975; Xu et al., 2006). Thus, we explicitly chose vowels because they more heavily tax perceptual-cognitive processing, and therefore listening effort, as might be revealed via pupillometry.

Tokens were 100 ms, including 10 ms of rise/fall time to reduce spectral splatter in the stimuli. Each contained identical voice fundamental (F0), second (F2), and third formant (F3) frequencies (F0: 150, F2: 1090, and F3: 2350 Hz). The F1 was parameterized over five equal steps between 430 and 730 Hz such that the resultant stimulus set spanned a perceptual phonetic continuum from /u/ to /a/ (Bidelman et al., 2013). Speech stimuli were delivered binaurally at 75 dB SPL through shielded insert earphones (ER-2; Etymotic Research) coupled to a TDT RP2 processor (Tucker Davis Technologies). This same speech continuum was presented in one of three noise blocks to vary SNR: unmasked, 0 dB SNR, −5 dB SNR. The masker was a speech-shaped noise based on the long-term power spectrum (LTPS) of the vowel set. While we typically use speech babble in our ERP studies, pilot testing showed this type of noise was too difficult for concurrent vowel identification, necessitating the use of simpler LTPS noise. The noise was presented continuously so that it was not time-locked to the stimulus presentation. Block order was randomized within and between participants.

During eyetracking, participants heard 150 trials of each speech token (per noise block). On each trial, participants labeled the sound with a binary response ("u" or "a") as quickly and accurately as possible. Following a behavioral response, the interstimulus interval (ISI) jittered randomly between 800 and 1000 ms (20 ms steps, uniform distribution) before the next trial commenced. EEG was also recorded during the categorization task. These data are reported elsewhere (Bidelman et al., 2019b).

Pupillometry Recording and Analysis

A Gazepoint GP3 eyetracker acquired listeners' gaze fixations based on published procedures from our laboratory (Bidelman et al., 2019a). This device provides precise measurement of the location of ocular gaze and pupil diameter with an accuracy of ∼1° visual angle via an infrared, desktop mounted camera. In addition to cognitive effort, a number of factors affect pupillometry including the pupillary light reflex (Fan and Yao, 2011) produced by the sympathetic nervous system (Andreassi, 2000). Consequently, the sound booth's lights remained off during the task. Participants could wear corrective lenses in the form of contacts. Continuous eye data were collected from the left and right eyes every 16.6 ms (i.e., 60 Hz sampling rate). MATLAB logged data from the GP3 via an API interface. Continued alignment with the screen was ensured by re-calibrating the eyetracker before each stimulus block. The GP3's internal routine calibrated the eyes at nine-points across the horizontal/vertical dimensions of the screen.

Continuous eye data were recorded online while participants performed the auditory CP task. A central fixation cross-hair (+) remained on the computer screen during the auditory task to center and maintain participants' gaze. Time stamps triggered in the data file demarcated the onset of each stimulus presentation. This allowed us to analyze time-locked changes in eye data for each stimulus akin to an evoked potential in the EEG literature (Beatty, 1982; Eckstein et al., 2017). Continuous recordings were filtered using a passband of 0.001–15 Hz, epoched [−100 to 1000 ms] (where t = 0 marks speech onset), baseline corrected, and ensemble averaged in the time domain to obtain the evoked pupil dilation response for each speech token per SNR and participant. This resulted in 15 waveforms per participant (= 5 tokens * 3 SNRs). Blinks were automatically logged by the eye tracker and epochs contaminated with these artifacts were discarded prior to analysis. Additionally, to correct for subtle changes in the distance between the eyetracker camera and the participant that could affect pupil measurements (e.g., during head movement), the Gazepoint records a continuous scale factor for each pupil; a scale value = 1 represents pupil depth (distance to the camera) at the time of calibration, scaling < 1 reflects when the user is closer to the eyetracker, and a scaling > 1 when the user is further away. This scale factor was then used to weight the running time course prior to averaging and correct for movement artifacts.

Data Analysis

Behavioral Data

Identification scores were fit with a sigmoid function $P = 1/[1 + e^{-\beta 1(x-\beta 0)}]$, where P is the proportion of trials

identified as a given vowel, x is the step number along the stimulus continuum, and β_0 and β_1 the location and slope of the logistic fit estimated using non-linear least-squares regression. Larger β_1 values reflect steeper psychometric functions and stronger categorical perception. Behavioral speech labeling speeds (i.e., reaction times; RTs) were computed as listeners' median response latency across trials for a given condition. RTs outside 250–2500 ms were deemed outliers (e.g., fast guesses, lapses of attention) and were excluded from analysis (Bidelman et al., 2013; Bidelman and Walker, 2017).

Pupillometry Data

To quantify the physiological data, we measured the peak (maximum) pupil diameter and latency within the search window between 300 and 700 ms. Visual inspection of the waveforms showed pupil responses were maximal in this timeframe (see **Figure 2**). Unless otherwise specified, dependent measures were analyzed using a two-way, mixed model ANOVA (subject = random factor) with fixed effects of SNR (three levels: unmasked, 0 and −5 dB SNR) and token [five levels: vw1-5] (PROC GLIMMIX, SAS® 9.4; SAS Institute, Inc.). Tukey–Kramer and Bonferroni adjustments were used to correct subsequent *post hoc* and planned multiple comparisons, respectively.

RESULTS

Behavioral Data

Bidelman et al. (2019b) fully describes the behavioral results. **Figure 1A** shows spectrograms of the individual speech tokens and **Figure 1B** shows behavioral identification functions across the SNRs. An analysis of slopes (β_1) revealed a main effect of SNR [$F_{2,28} = 35.25$, $p < 0.0001$] (**Figure 1C**). *Post hoc* contrasts confirmed that while 0 dB SNR did not alter psychometric slopes relative to unmasked speech ($p = 0.33$), the psychometric function became shallower with −5 dB SNR relative to 0 dB SNR ($p < 0.0001$). Additionally, SNR marginally but significantly shifted the perceptual boundary [$F_{2,28} = 5.62$, $p = 0.0089$] (**Figure 1D**). Relative to unmasked speech, −5 dB SNR speech shifted the perceptual boundary rightward ($p = 0.011$), suggesting a small but measurable bias to report "u" (i.e., more frequent vw1-2 responses) when noise exceeds the signal. Collectively, these results suggest that categorical representations are largely resistant to acoustic interference until signal strength of noise exceeds that of speech.

Behavioral response times (RTs) show the speed of categorization (**Figure 1E**). RTs varied with SNR [$F_{2,200} = 11.90$, $p < 0.0001$] and token [$F_{4,200} = 5.36$, $p = 0.0004$]. RTs were similar for unmasked and 0 dB SNR speech ($p = 1.0$) but slower for −5 dB SNR ($p < 0.0001$). *A priori* contrasts revealed this slowing was most prominent for more categorical tokens (vw1-2 and vw4-5). Ambiguous tokens (vw3) elicited similar RTs across noise conditions ($ps > 0.69$), suggesting that noise effects on RT were largely restricted to accessing categorical representations, not general slowing of decision speed across the board. We examined whether conditions elicited customary slowing in RTs near the midpoint of the continuum (Pisoni and Tash, 1974; Poeppel et al., 2004; Bidelman et al., 2013). Planned contrasts revealed this CP hallmark for unmasked [mean(vw1,2,4,5) vs. vw3; $p = 0.0003$] and 0 dB SNR ($p = 0.0061$) conditions, but not at −5 dB SNR ($p = 0.59$).

Pupillometry Data

Figure 2 shows grand average pupil waveforms for each speech token and SNR as well as the responses specifically contrasting unambiguous [mean (vw1,vw5)] vs. ambiguous (vw3) tokens. Visually, the data indicated that both SNR and the categorical status of speech modulated pupil responses. To quantify these effects, we pooled the peak (maximum) pupil diameter and latency of unambiguous tokens (vw1 and vw5) (those with stronger category identities) and compared them with the ambiguous vw3 token (Liebenthal et al., 2010; Bidelman, 2015; Bidelman and Walker, 2017). **Figure 3** shows the mean peak pupil diameters and latencies by SNR and behavioral RTs.

An ANOVA revealed a sole main effect of SNR on peak pupil size [$F_{2,196} = 6.69$, $p = 0.0015$] with no token [$F_{4,196} = 0.53$, $p = 0.7157$] nor token*SNR interaction effect [$F_{8,196} = 0.16$, $p = 0.9959$] (**Figure 3A**). Planned contrasts of pupil size between pairwise SNRs showed that only unmasked speech differed from intermediate SNR speech. Specifically, pupil diameter increased when classifying speech in moderate interference (i.e., 0 dB > unmasked; $p = 0.0007$) but did not differ with further increases in noise level (i.e., 0 dB = −5 dB; $p = 0.0794$) (**Figure 3B**).

An ANOVA on pupil latency revealed that SNR strongly modulated pupil response timing [$F_{2,196} = 4.60$, $p = 0.0112$], as did whether the token was unambiguous [$F_{4,196} = 3.25$, $p = 0.0130$] (**Figures 3C,D**). There was not a token*SNR interaction effect [$F_{8,196} = 0.94$, $p = 0.4827$]. Follow-up contrasts revealed similar latencies for unmasked and 0 dB speech ($p = 0.5379$), but longer latencies at −5 dB relative to 0 dB speech ($p = 0.0061$). Paralleling the RT data, *a priori* contrasts revealed an "inverted V-shaped" pattern analogous to the behavioral data—a slowing in response timing for ambiguous relative to unambiguous tokens in the 0 dB SNR [mean(vw1,2,4,5) vs. vw3; $p = 0.0244$]. Unmasked and −5 dB speech did not exhibit this pattern ($ps > 0.27$).

To further test whether behavior modulated eye behavior, we analyzed each listener's single-trial vw3 pupil responses based on (i) a median split of their behavioral RTs into fast and slow responses (**Figures 4A–E**) and (ii) the vowel category they reported (e.g., "a" vs. "u") (**Figures 4F–J**). This resulted in ∼75 trials for each subaverage. Despite having been elicited by an identical (though perceptually bistable) acoustic stimulus, vw3 pupil latencies were strongly dependent on the speed of listeners' decision [$F_{1,70} = 6.74$, $p = 0.0115$]. Slow RTs were associated with slower pupil responses to the ambiguous token (**Figure 4E**). Pupil size was not dependent on RTs [SNR, speed, and SNR × speed effects: $ps \geq 0.0585$] (**Figure 4D**). Split by listeners' identification (i.e., vw3 reported as "u" vs. "a"), we found a sole main effect of SNR on pupil response magnitudes [$F_{2,70} = 3.78$, $p = 0.0275$]. Pupil responses were again largest for 0 dB SNR speech compared to the other noise conditions (**Figure 4I**). These data reveal that under similar states of speech ambiguity, pupil responses

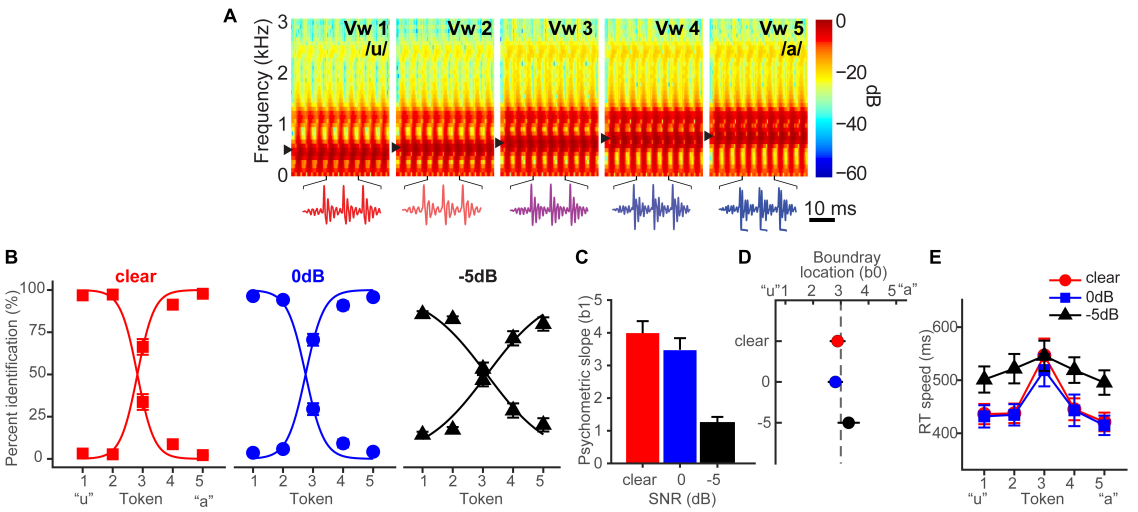

FIGURE 1 | Spectrograms and behavioral speech categorization at three levels of signal-to-noise ratio (SNR). **(A)** Spectrograms of individual speech tokens. **(B)** Perceptual psychometric functions. Note the curves are mirror symmetric reflecting the percentage of "u" (left curve) and "a" identification (right curve), respectively. **(C)** Slopes and **(D)** locations of the perceptual boundary show that speech categorizing is robust even down to 0 dB SNR. **(E)** Speech classification speeds (RTs) show a categorical slowing in labeling (Pisoni and Tash, 1974; Bidelman and Walker, 2017) for ambiguous tokens (midpoint) relative to unambiguous ones (endpoints) in unmasked and 0 dB SNR conditions. Categorization accuracy and speed deteriorate with noise interference by remains possible until severely degraded SNRs. Data reproduced from Bidelman et al. (2019b). Spectrogram reproduced from Bidelman et al. (2014), with permission from John Wiley & Sons. errorbars = ± SEM.

are modulated according to the speed of listeners' behavioral categorization. Note, this contrasts EEG findings for the same stimuli, which show that electrical brain activity differentiates the ambiguous speech depending on listeners' subjective report (i.e., vw3 heard as "u" vs. "a") (Bidelman et al., 2013).

DISCUSSION

By recording continuous pupil responses during a rapid speech categorization task in noise, we assessed how acoustic interference impacts cognitive load and perceptual identification of phonemes. Our analyses revealed that speech perception was robust to moderate acoustic interference (i.e., $\geq$ 0 dB SNR). More category representative (less ambiguous) phonetic tokens reduced listening effort and were more resilient to moderate acoustic interference. While noise impacts perception of ambiguous phonemes, categorical coding appears to mitigate interference by enhancing representations of phonemes. We propose that categorical coding (i.e., speech with an unambiguous identity) helps partially counteract the negative effects of noise on perception, but only to the extent that speech signals are not too severely degraded. Our findings converge with notions that the process of categorization aids the extraction of speech from noise whereby abstract categories help fortify the speech code and make it more resistant to external noise interference (e.g., Helie, 2017; Bidelman et al., 2019b).

Physiologically, our data suggest that difficulty of speech processing modulates pupil behavior, but not straightforwardly. It is a common finding that pupil size increases when tasks are difficult to perform (Beatty, 1982). Consistent with our predictions, pupil size increased for moderately corrupted relative to unmasked speech but plateaued for severely corrupted speech. Previous work has assessed the pupil response to speech (sentences) across a broad range of intelligibility levels [i.e., −36 to −4 dB in nine 4 dB steps] (Zekveld and Kramer, 2014). This work suggests that pupil dilation increases at intermediate SNRs, but minimally at low and high SNRs, which has been interpreted to reflect intelligibility and/or task difficulty (Ohlenforst et al., 2017). The fact that pupil diameter of our participants increased with moderate SNR suggests the task demands in this condition did not exceed available cognitive resources. A recent pupillometry study found that pupil behavior correlates with subjective ratings of salience defined in terms of how noticeable or remarkable sounds are considered, indicating greater listening demand or arousal (Liao et al., 2016). In this vein, our result might reflect a performance/arousal tradeoff known as Yerkes-Dodson law, a phenomenon where performance resembles an inverted-U function of arousal (Yerkes and Dodson, 1908). Pupil dilation correlates with arousal responses measured in the locus coeruleus (LC) (Aston-Jones and Cohen, 2005). A variety of cognitive tasks elicit a strong relationship between performance and LC activity, whereby activation in the middle of the Yerkes-Dodson curve is associated with increased performance and task engagement (for reviews, see Berridge and Waterhouse, 2003; Aston-Jones and Cohen, 2005; Sara and Bouret, 2012). Under this framework, listeners are less attentive and disengaged (hypoarousal) and thus perform more poorly; when LC activity increases beyond intermediate range, listeners would be more distracted (hyperarousal), which would also reduce performance. Interestingly, a neuroimaging study reported a similar finding in neural responses over left

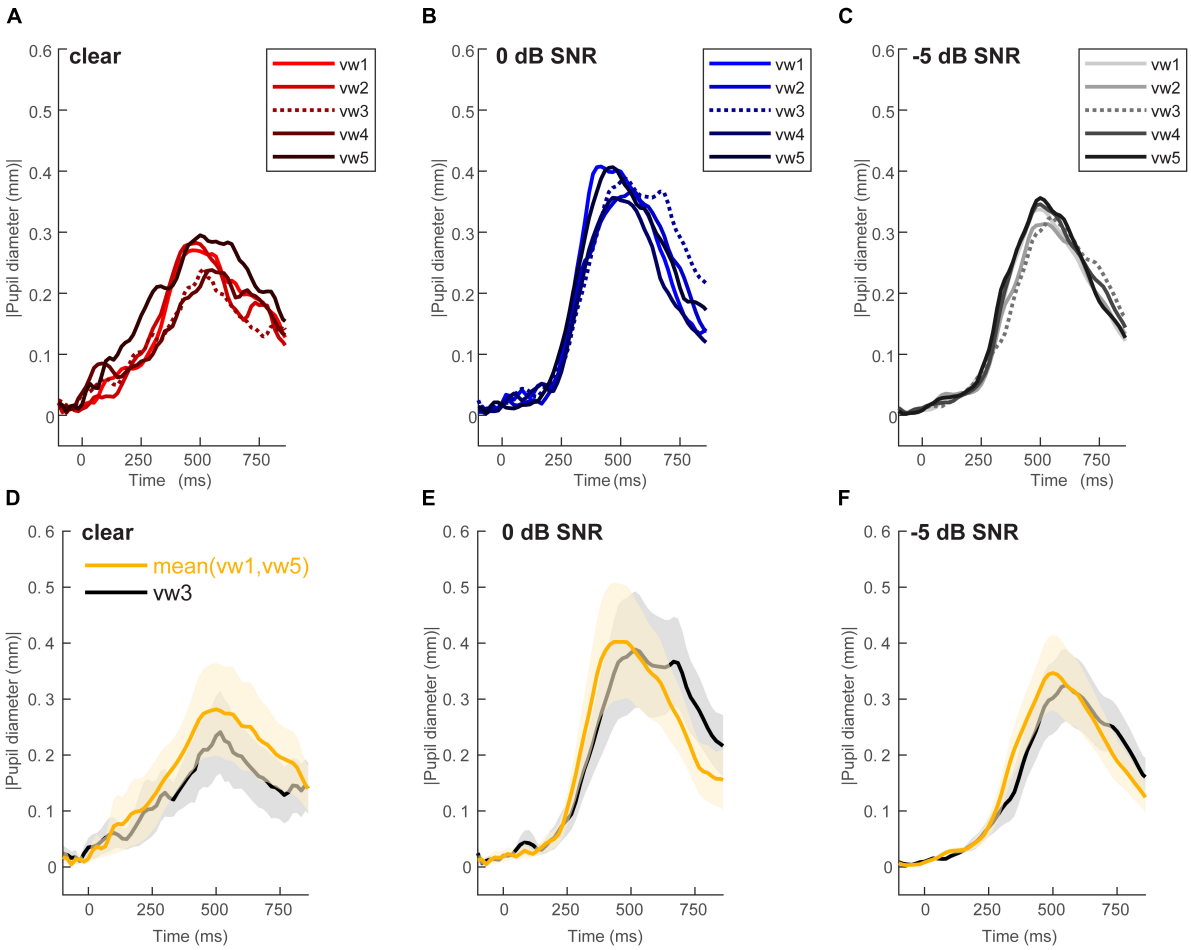

FIGURE 2 | Grand average waveforms for pupil responses. Average responses to each token condition at each SNR level: **(A)** unmasked, **(B)** 0 dB SNR, **(C)** –5 dB SNR conditions. Peak pupil diameter and latency between the 300 and 700 ms search window are extracted for further analysis. Grand average waveforms for pupil responses contrasting categorical [mean (vw1,vw5)] vs. ambiguous (vw3) tokens at each SNR level. **(D)** Unmasked, **(E)** 0 dB SNR, **(F)** –5 dB SNR conditions. Pupil responses are modulated by SNR and token identity. shading = 1 SEM.

temporal cortex and premotor cortex, with greater activity for slightly degraded speech relative to unmasked and severely degraded speech (Davis and Johnsrude, 2003), paralleling our pupillometry results.

The most interesting findings were for pupil latency. Previous work has shown that reduced speech intelligibility systematically delays pupil responses (Zekveld et al., 2010), implying increased listening effort. While we found responses were more delayed at severe than intermediate noise levels, latencies for unmasked and intermediate speech did not differ overall (i.e., unmasked = 0 dB). Listeners may have compensated by exerting more effort in the intermediate noise condition (McGarrigle et al., 2017). Importantly, pupil responses were more categorical at intermediate SNRs, as evidenced by a slowing in pupil responses for ambiguous tokens. This pattern was not observed at −5 dB SNR. These findings suggest categorical coding helps reconstruct degraded speech sounds with unambiguous identities, but only within a limited range of intelligibility.

Behaviorally, psychometric slopes were steeper for unmasked relative to noise-degraded speech, and only became flatter for severely degraded speech. Indeed, only highly degraded speech weakened CP, further suggesting that the natural binning process of categorical coding helps maintain robust perception of SIN. Presumably, CP enhances processing within the acoustic space to help phonetic representations stand out (e.g., Nothdurft, 1991; Perez-Gay et al., 2018). We argue that noise-related decrements in CP reflect weakening of internalized categories rather than less vigilant listening across the board because ambiguous tokens elicited similar RTs across noise levels. Moreover, both our behavioral and physiological data indicated more categorical responses to unambiguous relative to ambiguous tokens at intermediate noise levels. Thus, noise-related decrements in our data likely reflect fuzzier matches between speech signals and templates of speech sounds (Bidelman et al., 2019b).

Discrepancies between the behavioral and physiological data in SNR which showed categorical coding (i.e., inverted-V pattern) suggest perhaps that pupil responses are less sensitive than

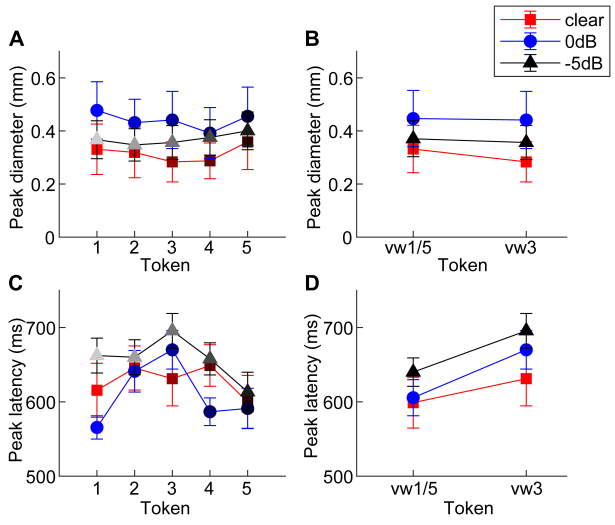

FIGURE 3 | Mean peak pupil diameters and latencies by SNR. **(A)** Larger pupil size is observed at 0 dB SNR relative to unmasked and –5 dB SNR. **(B)** Peak pupil diameter is elevated at 0 dB SNR relative to the other two conditions. **(C,D)** In general, –5 dB speech shows the longest peak latencies of the three conditions. Pupil responses are delayed for 0 dB SNR speech and for categorically ambiguous speech (i.e., vw3 > vw1/5). errorbars = 1 SEM.

behavior and require the additional "load" of intermediate noise to show a categorical effect in response timing. Additionally, while the −5 dB condition produced significantly worse behavioral performance relative to quiet, it was the 0 dB condition instead that produced larger peak pupil dilation. This could reflect the fact that the 0 dB condition was more effortful than quiet, despite behavioral accuracy remaining high. Such findings align with notions of the FUEL model (Pichora-Fuller et al., 2016) suggesting performance is governed by a combination of signal quality (e.g., input SNR) and internal factors (e.g., arousal, attention, and motivation).

One interpretation of CP is that ambiguous or intermediate tokens are "drawn" toward prototypes or category centers, i.e., the veridical percept is warped by the existence of a category representation such that peripheral tokens are perceived as more central (e.g., "perceptual magnet" theory; Kuhl, 1991; Iverson et al., 2003). Our physiological data loosely align with this notion, showing and influence of category prototypicality/centrality on degraded speech perception. Peripheral tokens (e.g., vw2 and vw4) elicited similar pupil responses to their central prototype (i.e., continuum endpoints), as evidenced by the inverted-V pattern in RT (**Figure 1E**) and pupil latency data (**Figure 3C**). Still, for speech sounds which split the perceptual boundary (i.e., vw3)—and are thus perceptually ambiguous—we find this perceptual draw is considerably weaker if made at all. This is supported by the fact pupil responses to vw3 were similar when split by listeners' subjective report ("u" vs. "a"; **Figures 4F–H**). Collectively, these later findings align with more relaxed models of perception which consider gradiency, whereby the system must balance the efficiency of discarding potential rich and continuous acoustic details with discrete category representations (McMurray et al., 2008). Thus, one might equally discuss our findings as reflecting the gradience of phonetic categories (especially vowels), and more generally perceptual uncertainty, rather than CP *per se*. Under this interpretation, acoustic cues that allow the rapid assessment of category membership of

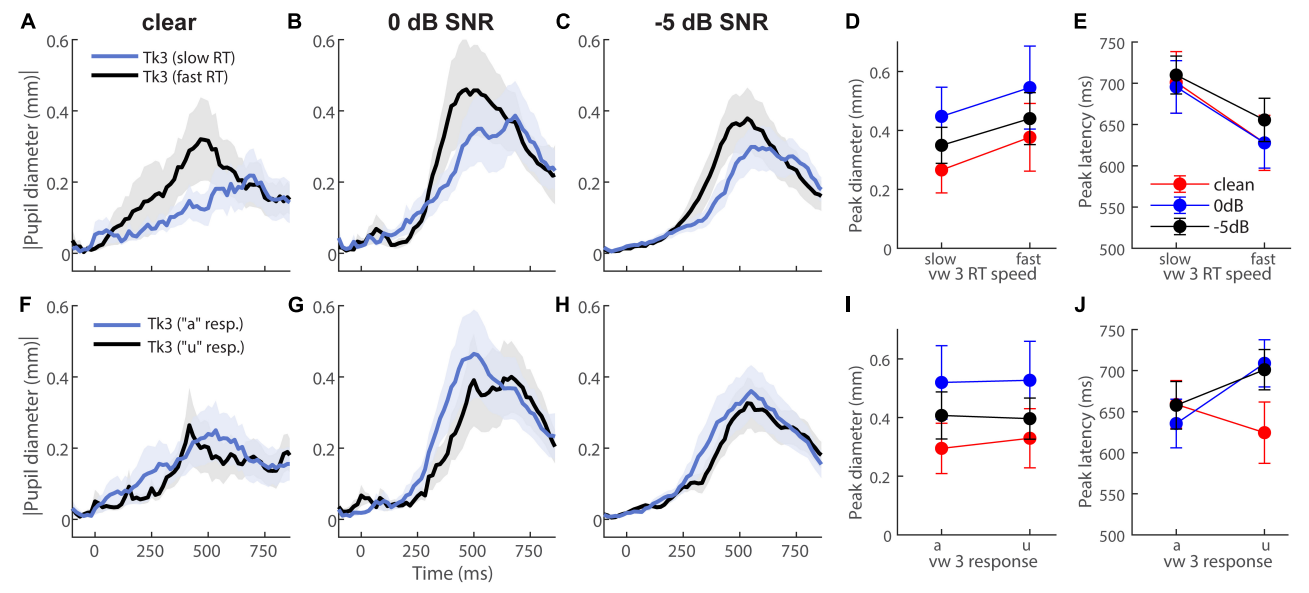

FIGURE 4 | Pupil response latency but not size depends on speed of listeners' decision. Grand average waveforms for pupil responses to vw3 based on **(A–E)** a median split of behavioral RTs and **(F–J)** the reported vowel category (e.g., "a" vs. "u"). **(E)** Pupil latencies strongly depend on speed of listeners' decision. Slow RTs are associated with slower pupil responses to ambiguous token. **(D)** Pupil size is not dependent on RTs. **(I)** SNR has a sole effect on pupil response magnitudes when split by listeners' identification (i.e., reporting vw3 as "u" vs. "a"). Pupil responses are again largest for 0 dB SNR speech compared to other noise conditions.

unambiguous tokens (e.g., vw1, vw5) are acoustically/perceptually available until noise masking is too egregious. In cases in which speech cues are ambiguous (vw3), noise fails to alter the decision process much, because listeners are already dealing with ambiguous acoustic-phonetic information.

Collectively, our findings converge with notions that categorical representations of phonemes are more salient and resilient to noise degradation than acoustic-sensory ones (Helie, 2017; Bidelman et al., 2019b,c). On the premise that phonetic representations (a high-level code) are more resilient to noise than surface level features (a low-level code) (Helie, 2017; Bidelman et al., 2019b,c), the construction of perceptual objects and natural binning process of CP might mitigate noise by helping category members stand out among a noisy feature space. Despite being acoustically dissimilar, categorically equivalent sounds would elicit similar changes in local firing rate, whereas cross-category (perceptually distinct) sounds would not (e.g., Recanzone et al., 1993; Guenther and Gjaja, 1996; Guenther et al., 2004). Noise would create a noisier map for physical acoustic details, but phonetic categories would persist (e.g., Nothdurft, 1991; Perez-Gay et al., 2018).

We found that ambiguous speech increased listening effort (delayed pupil responses). Results from fMRI similarly suggest that activation of auditory cortical cells may be shorter for category prototypes than for other sounds (Guenther et al., 2004). Indeed, participants labeled unambiguous tokens more quickly than ambiguous tokens, suggesting more efficient processing of members from well-formed categories. This advantage was also observed in pupil latencies in the intermediate noise condition, but not in the unmasked condition. Delayed pupil responses might instead reflect processes of ambiguity resolution. In speech, there is no one-to-one correspondence between any single acoustic cue and phonetic representations (Lotto and Holt, 2016). Partial loss of acoustic cues would render phonemes highly confusable with one another. Connectionist models of speech perception such as TRACE (McClelland and Elman, 1986) posit bi-directional, interactive activation of phonemic traces that help recover meaning when signal features are missed. Under TRACE, speech processing transpires through a neuronal network representing speech features at increasingly higher levels. Incoming acoustic input activates nodes for features (and inhibits others), which in turn activate phonemes at the next level. During this process, traces of inhibited representations remain activated for a period, helping the listener recover information if errors are perceived (e.g., missing an acoustic segment). If noise leads to partial loss of cues, delayed pupil responses observed in our data might reflect ongoing activation (through a TRACE-like network) of multiple phonetic representations in attempt to disambiguate what is being heard.

In sum, the present findings demonstrate that pupillometry can be used as an effective technique for assessing underlying processes of speech perception and categorical processing. Here, the benefits of tracking CP with pupillometry were twofold: (a) providing complementary physiological data for comparison with existing data, and (b) lending temporally sensitive insight into mental processes not available from behavioral measures alone.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The University of Memphis IRB. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award number R01DC016267 (GB).

ACKNOWLEDGMENTS

We thank Lauren Bush and Alex Boudreaux for assistance in data collection.

REFERENCES

Altmann, C. F., Uesaki, M., Ono, K., Matsuhashi, M., Mima, T., and Fukuyama, H. (2014). Categorical speech perception during active discrimination of consonants and vowels. *Neuropsychologia* 64C, 13–23. doi: 10.1016/j.neuropsychologia.2014.09.006

Andreassi, J. L. (ed.). (2000). "Pupillary Response and Behavior," in *Psychophysiology: Human Behavior & Physiological Response*. New Jersey: Lawrence Erbaulm Associates, Inc, 289–307.

Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* 91, 276–292. doi: 10.1037//0033-2909.91.2.276

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, M. K., Reingold, E. M., and Schneider, B. A. (2011). Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements. *J. Speech, Lang. Hea. Res.* 54, 243–262. doi: 10.1044/1092-4388(2010/09-0233)

Berridge, C. W., and Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Res. Rev.* 42, 33–84. doi: 10.1016/s0165-0173(03)00143-7

Bidelman, G. M. (2015). Induced neural beta oscillations predict categorical speech perception abilities. *Brain Lang.* 141, 62–69. doi: 10.1016/j.bandl.2014.11.003

Bidelman, G. M., Brown, B., Mankel, K., and Price, C. N. (2019a). Psychobiological responses reveal audiovisual noise differentially challenges speech recognition. *Ear. Hear.* doi: 10.1097/AUD.0000000000000755 [Epub ahead of print].

Bidelman, G. M., Bush, L. C., and Boudreaux, A. M. (2019b). The categorical neural organization of speech aids its perception in noise. *bioRxiv* [preprint]. doi: 10.1101/652842

Bidelman, G. M., Sigley, L., and Lewis, G. (2019c). Acoustic noise and vision differentially warp speech categorization. *J. Acoust. Soc. Am.* 146, 60–70. doi: 10.1121/1.5114822

Bidelman, G. M., and Lee, C.-C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *Neuroimage* 120, 191–200. doi: 10.1016/j.neuroimage.2015.06.087

Bidelman, G. M., Moreno, S., and Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* 79, 201–212. doi: 10.1016/j.neuroimage.2013.04.093

Bidelman, G. M., and Walker, B. (2017). Attentional modulation and domain specificity underlying the neural organization of auditory categorical perception. *Eur. J. Neurosci.* 45, 690–699. doi: 10.1111/ejn.13526

Bidelman, G. M., Weiss, M. W., Moreno, S., and Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2673. doi: 10.1111/ejn.12627

Bidelman, G. M. (2017). "Communicating in challenging environments: noise and reverberation," in *Springer Handbook of Auditory Research: The frequency-following response: A window into human communication. vol Springer Handbook of Auditory Research*, Vol. 61, eds N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, and A. N. Popper (New York, N.Y: Springer Nature).

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., and Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nat. Neurosci.* 7, 295–301. doi: 10.1038/nn1198

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* 13, 1428–1432. doi: 10.1038/nn.2641

Cousins, K. A., Dar, H., Wingfield, A., and Miller, P. (2014). Acoustic masking disrupts time-dependent mechanisms of memory encoding in word-list recall. *Mem. Cogn.* 42, 622–638. doi: 10.3758/s13421-013-0377-7

Davis, M. H., and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J. Neurosci.* 23, 3423–3431. doi: 10.1523/jneurosci.23-08-03423.2003

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

Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., and Scott, S. K. (2016). Getting the cocktail party started: masking effects in speech perception. *J. Cogn. Neurosci.* 28, 483–500. doi: 10.1162/jocn_a_00913

Fan, X., and Yao, G. (2011). Modeling transient pupillary light reflex induced by a short light flash. *IEEE Trans. Biomed. Eng.* 58, 36–42. doi: 10.1109/TBME.2010.2080678

Gatehouse, S., and Gordon, J. (1990). Response times to speech stimuli as measures of benefit from amplification. *Br. J. Audiol.* 24, 63–68. doi: 10.3109/03005369009077843

Goldstone, R. L., and Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdiscip. Rev.Cogn. Sci.* 1, 69–78. doi: 10.1002/wcs.26

Guenther, F. H., and Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *J. Acoust. Soc. Am.* 100(2 Pt 1), 1111–1121. doi: 10.1121/1.416296

Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., and Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *J. Speech Lang. Hear Res.* 47, 46–57. doi: 10.1044/1092-4388(2004/005)

Harnad, S. R. (1987). *Categorical Perception: The Groundwork of Cognition*. New York, NY: Cambridge University Press.

Heinrich, A., Schneider, B. A., and Craik, F. I. (2008). Investigating the influence of continuous babble on auditory short-term memory performance. *Q. J. Exp. Psychol.* 61, 735–751. doi: 10.1080/17470210701402372

Helie, S. (2017). The effect of integration masking on visual processing in perceptual categorization. *Brain Cogn.* 116, 63–70. doi: 10.1016/j.bandc.2017.06.001

Hyönä, J., Tommola, J., and Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Q. J. Exp. Psychol.* 48, 598–612. doi: 10.1080/14640749508401407

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47–B57.

Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science* 154, 1583–1585. doi: 10.1126/science.154.3756.1583

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 116(4 Pt 1), 2395–2405. doi: 10.1121/1.1784440

Koelewijn, T., Shinn-Cunningham, B. G., Zekveld, A. A., and Kramer, S. E. (2014). The pupil response is sensitive to divided attention during speech processing. *Hear. Res.* 312, 114–120. doi: 10.1016/j.heares.2014.03.010

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* 50, 93–107. doi: 10.3758/bf03212211

Liao, H.-I., Kidani, S., Yoneya, M., Kashino, M., and Furukawa, S. (2016). Correspondences among pupillary dilation response, subjective salience of sounds, and loudness. *Psychon. B Rev.* 23, 412–425. doi: 10.3758/s13423-015-0898-0

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

Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., and Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cereb. Cortex* 20, 2958–2970. doi: 10.1093/cercor/bhq045

Livingston, K. R., Andrews, J. K., and Harnad, S. (1998). Categorical perception effects induced by category learning. *J. Exp. Psychol.Learn. Mem. Cogn.* 24, 732–753. doi: 10.1037//0278-7393.24.3.732

López-Ornat, S., Karousou, A., Gallego, C., Martín, L., and Camero, R. (2018). Pupillary measures of the cognitive effort in auditory novel word processing and short-term retention. *Front. Psychol.* 9:2248. doi: 10.3389/fpsyg.2018.02248

Lotto, A. J., and Holt, L. L. (2016). Speech perception: the view from the auditory system. in *Neurobiology of Language* eds G. Hickok, and S. L. Small. (Amsterdam:: Elsevier), 185–194.

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cognit. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., and Munro, K. J. (2017). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *J. Exp. Child Psychol.* 161, 95–112. doi: 10.1016/j.jecp.2017.04.006

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., and Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1609–1631. doi: 10.1037/a0011747

Myers, E. B., and Swan, K. (2012). Effects of category learning on neural sensitivity to non-native phonetic categories. *J. Cogn. Neurosci.* 24, 1695–1708. doi: 10.1162/jocn_a_00243

Naylor, G., Koelewijn, T., Zekveld, A. A., and Kramer, S. E. (2018). *The Application of Pupillometry in Hearing Science to Assess Listening Effort*. Los Angeles, CA: SAGE Publications Sage CA.

Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Res.* 31, 1073–1078. doi: 10.1016/0042-6989(91)90211-m

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

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Perez-Gay, F., Sicotte, T., Theriault, C., and Harnad, S. (2018). Category learning can alter perception and its neural correlate. *arXiv* [Preprint]. Available at: https://arxiv.org/abs/1805.04619 (accessed November 4, 2019).

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: the

framework for understanding effortful listening (FUEL). *Ear. Hear.* 37, 5S–27S. doi: 10.1097/AUD.0000000000000312

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept. Psychophys.* 13, 253–260. doi: 10.3758/bf03214136

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Mem. Cognit.* 3, 7–18. doi: 10.3758/BF03198202

Pisoni, D. B., and Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition* 25, 21–52. doi: 10.1016/0010-0277(87)90003-5

Pisoni, D. B., and Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Percept. Psychophys.* 15, 285–290. doi: 10.3758/bf03213946

Poeppel, D., Guillemin, A., Thompson, J., Fritz, J., Bavelier, D., and Braun, A. R. (2004). Auditory lexical decision, categorical perception, and FM direction discrimination differentially engage left and right auditory cortex. *Neuropsychologia* 42, 183–200. doi: 10.1016/j.neuropsychologia.2003.07.010

Recanzone, G. H., Schreiner, C. E., and Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* 13, 87–103. doi: 10.1523/jneurosci.13-01-00087.1993

Renner, L. F., and Wlodarczak, M. (2017). *When a Dog is a Cat and How it Changes Your Pupil Size: Pupil Dilation in Response to Information Mismatch*. Stockholm: INTERSPEECH, 2017, 674–678.

Rönnberg , J., Lunner, T., Zekveld, A., Sörqvist , P., Danielsson, H., Lyxell, B., et al. (2013). The Ease of Language Understanding (ELU) model: theory, data, and clinical implications. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031

Sara, S. J., and Bouret, S. (2012). Orienting and reorienting: the locus coeruleus mediates cognition through arousal. *Neuron* 76, 130–141. doi: 10.1016/j.neuron.2012.09.011

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406. doi: 10.1093/brain/123.12.2400

Sharma, A., and Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J. Acoust. Soc. Am.* 106, 1078–1083. doi: 10.1121/1.428048

Vogelzang, M., Hendriks, P., and van Rijn, H. (2016). Pupillary responses reflect ambiguity resolution in pronoun processing. *Lang. Cogn. Neurosci.* 31, 876–885. doi: 10.1080/23273798.2016.1155718

Wendt, D., Dau, T., and Hjortkjaer, 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. 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., 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 Hea.* 22:2331216518800869. doi: 10.1177/2331216518800869

Xu, Y., Gandour, J. T., and Francis, A. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J. Acoust. Soc. Am.* 120, 1063–1074. doi: 10.1121/1.2213572

Yerkes, R. M., and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit—formation. *J. Comp. Neurol. Psychol.* 18, 459–482. doi: 10.1002/cne.920180503

Yoo, J., and Bidelman, G. M. (2019). Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception. *Hear. Res.* 377, 189–195. doi: 10.1016/j.heares.2019.03.021

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

Zekveld, A. A., and Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: insights from pupillometry. *Psychophysiology* 51, 277–284. doi: 10.1111/psyp.12151

Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2010). Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear. Hear.* 31, 480–490. doi: 10.1097/AUD.0b013e3181d4f251

Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2011). Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. *Ear. Hear.* 32, 498–510. doi: 10.1097/AUD.0b013e31820512bb

Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., and Li, P. (2011). Cortical dynamics of acoustic and phonological processing in speech perception. *PLoS One* 6:e20963. doi: 10.1371/journal.pone.0020963

Conflict of Interest: 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.