NeuroImage 269 (2023) 119899

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Perceptual warping exposes categorical representations for speech in human brainstem responses

Jared A. Carter[a,b,c], Gavin M. Bidelman[d,e,*]

[a] Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
[b] School of Communication Sciences and Disorders, University of Memphis, Memphis, TN, USA
[c] Division of Clinical Neuroscience, School of Medicine, Hearing Sciences – Scottish Section, University of Nottingham, Glasgow, Scotland, UK
[d] Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, IN, USA
[e] Program in Neuroscience, Indiana University, Bloomington, IN, USA

ARTICLE INFO

Keywords:
Frequency following response (FFR)
Categorical perception (CP)
Nonlinear dynamics
Hysteresis

ABSTRACT

The brain transforms continuous acoustic events into discrete category representations to downsample the speech signal for our perceptual-cognitive systems. Such phonetic categories are highly malleable, and their percepts can change depending on surrounding stimulus context. Previous work suggests these acoustic-phonetic mapping and perceptual warping of speech emerge in the brain no earlier than auditory cortex. Here, we examined whether these auditory-category phenomena inherent to speech perception occur even earlier in the human brain, at the level of auditory brainstem. We recorded speech-evoked frequency following responses (FFRs) during a task designed to induce more/less warping of listeners' perceptual categories depending on stimulus presentation order of a speech continuum (random, forward, backward directions). We used a novel clustered stimulus paradigm to rapidly record the high trial counts needed for FFRs concurrent with active behavioral tasks. We found serial stimulus order caused perceptual shifts (hysteresis) near listeners' category boundary confirming identical speech tokens are perceived differentially depending on stimulus context. Critically, we further show neural FFRs during active (but not passive) listening are enhanced for prototypical vs. category-ambiguous tokens and are biased in the direction of listeners' phonetic label even for acoustically-identical speech stimuli. These findings were not observed in the stimulus acoustics nor model FFR responses generated via a computational model of cochlear and auditory nerve transduction, confirming a central origin to the effects. Our data reveal FFRs carry category-level information and suggest top-down processing actively shapes the neural encoding and categorization of speech at subcortical levels. These findings suggest the acoustic-phonetic mapping and perceptual warping in speech perception occur surprisingly early along the auditory neuroaxis, which might aid understanding by reducing ambiguity inherent to the speech signal.

1. Introduction

To effectively utilize speech, individuals must convert continuous stimuli in the external world to phonetic category units (Goldstone and Hendrickson, 2010). In continuous speech, the precise acoustic characteristics of phonemes vary depending on the speaker (e.g., sex, accent) (Sumner, 2011), surrounding coarticulation (Beddor et al., 2002), and background noise (Bidelman, 2016; Billings et al., 2009; Carter and Bidelman, 2021). Categorization allows this variation to exist without hindering the utility of speech as a mode of communication. One open question in categorization is whether its driving force lies in neurophysiological constraints of the sensory system (i.e., bottom-up coding of sound) (Kuhl, 1986; Kuhl and Miller, 1975) or if higher-order language and memory regions modulate categorical speech percepts in a top-down manner (Bidelman et al., 2021; Carter and Bidelman, 2021; Ganong and Zatorre, 1980; Kuhl, 1986; Kuhl and Miller, 1975). If top-down modulations of early speech representations do occur, then how far down the auditory system are these perceptual influences exerted?

Typically, when assessing categorization, signals are presented to listeners who are asked to identify the sound as a member of a set of discrete categories. Their behavioral responses can be represented as a psychometric function, which can be quantified by its slope and its categorical boundary. A steeper slope indicates the perceptual change from one category to the next happens more rapidly than if the slope was shallower and thus indexes the strength of categorical hearing across the continuum (Bidelman, 2015a; Strouse et al., 1998; Xu et al., 2006a). The categorical boundary indicates the point at which the psychometric function crosses 50% identification, marking the stimulus location

* Corresponding author at: Speech, Language and Hearing Sciences, 2631 East Discovery Parkway, Bloomington, IN 47408, USA.
E-mail address: gbidel@indiana.edu (G.M. Bidelman).

https://doi.org/10.1016/j.neuroimage.2023.119899.
Received 13 July 2022; Received in revised form 17 January 2023; Accepted 22 January 2023
Available online 28 January 2023.

where the category shifts from one percept to another (Altmann et al., 2014; Ganong III and Zatorre, 1980). Additionally, one can measure how rapidly a listener labels each token via reaction time (RT). RTs demonstrate the speed of processing, which increases (i.e., slows down) during more ambiguous or degraded tokens and decreases (i.e., speeds up) during more prototypical tokens, often yielding an inverted U shape when plotting RTs across the continua (Pisoni and Tash, 1974).

The categorical perception of speech is usually assessed by randomizing the presentation of stimuli from a graded acoustic continuum. When presenting stimuli in sequential order (e.g., high-to-low first formant frequency [F1]), rather than a random order, the categorical boundary is modified due to short-term sequencing effects (Diehl et al., 1978; Healy and Repp, 1982). Such perceptual shifts may reflect a perseveration of the prior perception (i.e., hysteresis) or changing perception to the other category earlier than anticipated (i.e., enhanced contrast) (Tuller et al., 1994). These types of dynamics in perception suggest the brain's ongoing sorting of incoming acoustics into categorical phonetic representations is actively modulated during perception. Whether warping is due to top-down (Bathellier et al., 2012; Carter et al., 2022; Tuller et al., 2008) vs. bottom-up (e.g., adaptation of "phonetic feature detectors") (Eimas and Corbit, 1973) mechanisms is debatable. Presumably, such effects are driven more by top-down processes since they are observed during active listening tasks. Though, the role of top-down vs. bottom-up processing in perceptual warping has not been formally investigated. We address these questions in the current study.

When viewed through the lens of nonlinear dynamic systems, this process can be described as a shifting of the perceptual space to accommodate variability within categories (Tuller et al., 1994). Such warpings in perceptual space are likely driven by prefrontal (i.e., memory) brain regions that track ongoing stimulus history and adjust current percepts according to listeners' expectations and perceptual biases (Carter et al., 2022; Hansen et al., 2006). We do not yet know how far down the auditory system this top-down modulation of speech representation continues, however. While fronto-temporal pathways drive auditory stimulus encoding in cortex, the corticofugal system (i.e., cortico-collicular efferent pathways) can also modulate responses in the auditory brainstem by fine-tuning sound representations according to listening demands (Suga, 2008; Suga et al., 2000). Additionally, corticofugal fibers enhance speech processing prior to its arrival in cortex through attention-dependent gain control (Lai et al., 2022; Price and Bidelman, 2021). This makes the corticofugal system a prime candidate for tuning speech representations and possibly building nascent acoustic-phonetic structure at *subcortical* levels.

The frequency-following response (FFR) has been used as a window to characterize early, subcortical sound encoding along the auditory system. The FFR is a scalp-recorded potential evoked by sustained stimuli (such as speech) occurring ~7-10 milliseconds after stimulus onset with putative source(s) in the auditory brainstem (i.e., inferior colliculus) (Bidelman, 2018b; Gardi et al., 1979; Langner and Schreiner, 1988; Smith et al., 1975; Sohmer et al., 1977), and not cochlear origin (Skoe and Kraus, 2010). Early work in animal models localized the FFR to several subcortical auditory nuclei including cochlear nucleus (CN), inferior colliculus (IC), and medial geniculate body (MGB) (Dunlop et al., 1965; Oatman and Anderson, 1980; Sohmer et al., 1977). While most previous work has shown a subcortical origin of the FFR, recent neuroimaging studies have suggested cortical contributions to the response at low (<100 Hz) frequencies when recorded via magnetoencephalography (MEG) (Coffey et al., 2016; Gorina-Careta et al., 2021; Ross et al., 2020; Tang et al., 2016) or intracortically (Gnanateja et al., 2021). Latency modeling also supports more cortical involvement for low-frequency stimuli (Tichko and Skoe, 2017), where cortical neurons are still able to robustly synchronize (Joris et al., 2004) before transformation to a rate-based representation at higher frequencies (starting as low as ~50 Hz; Tang et al., 2016). Cortical contributions to the FFR may be a contributing and/or modulatory factor of the overall response, with stimulus frequencies and recording factors biasing more/less involvement of cortex to the scalp-recorded response (Bidelman, 2018b; Coffey et al., 2019). On the contrary, EEG work has convincingly demonstrated that subcortical structures (i.e., midbrain and even auditory nerve) provide the largest contribution to the scalp-recorded FFR_{EEG} for most of the frequency bandwidth of speech (Bidelman, 2018b; Bidelman and Momtaz, 2021). The FFR phase-locks with the time-varying, spectro-temporal features of complex sounds including fundamental frequency (F0) and harmonics (Galbraith et al., 1995), as well as the first few formant frequencies up to its phase locking limits (~1200 Hz) (Aiken and Picton, 2008; Krishnan, 2002; Skoe and Kraus, 2010). Given its unique time-frequency signature within the EEG, FFRs have been used to characterize subcortical processing of speech (Bidelman and Powers, 2018; Bidelman and Momtaz, 2021; Bidelman et al., 2013; Galbraith et al., 1995; Johnson et al., 2005; Musacchia et al., 2008; Russo et al., 2004; Skoe and Kraus, 2010) and musical sounds (Bidelman, 2013; Bones et al., 2014; Mankel and Bidelman, 2018), as well as track changes in neural encoding across the lifespan (Anderson et al., 2012; Bidelman et al., 2019, 2014b; Liu et al., 2018). Of interest for this study is the use of FFRs in understanding the brain's earliest neural representations for speech and its sensitivity to specific phonetic features found in a listeners' native language (cf. categories) (Krishnan et al., 2010; Krishnan et al., 2009).

To date, categorical representations have not been observed in brainstem FFRs, which, despite their ability to faithfully encode speech stimulus properties (e.g., formants), do not show strong evidence of category structure. Using an active categorization task (/u/ to /ɑ/ continuum), Bidelman et al. (2013) found that neurometric identification functions derived from the auditory *cortical* ERPs (~175 ms) predicted listeners' behavioral psychometric identification functions. In contrast, similar neurometric functions derived from *brainstem* FFRs did not. The data suggested that while category representations are observed at a cortical level, brainstem is perhaps too early along the auditory processing hierarchy to observe abstract category structure. However, an important caveat of this study was that FFRs, as in most subcortical studies, were recorded under passive listening tasks. Indeed, Bidelman and Walker (2017) demonstrated that categorical representations only manifest with goal-directed attention and under active (but not passive) speech identification tasks. Such attention-dependent effects might arise due to engagement of a wider network of category-sensitive brain regions and reciprocal connections between inferior frontal and auditory cortical areas which guide the formation of perceptual objects (Alho et al., 2016; Carter and Bidelman, 2021; Carter et al., 2022).

Nevertheless, some evidence shows category representations might exist in subcortical structures. In guinea pig, auditory brainstem responses evoked by noise bursts separate in a nonlinear fashion (indicative of categorical coding) based on the gap duration between noise bursts (Burghard et al., 2019). Studies that compared listeners fluent in tonal (e.g., Chinese) vs. non-tonal (e.g., English) languages show that the former tend to have stronger pitch representation in subcortical responses, but only for pitches that match native pitch contours in their language (Krishnan et al., 2010; Krishnan et al., 2009; Xu et al., 2006a). However, this effect does not carry over to similar acoustic analogues of the pitch changes that are not found in the native tone space (Xu et al., 2006b). Such findings suggest the presence of linguistically-relevant (categorical-like) information in the brainstem, but itself does not indicate the active process of categorization is occurring locally in brainstem, *per se*. Such findings could be explained by long term, experience-dependent plasticity (Krishnan et al., 2012). This evidence is further bolstered by findings of categorization-training studies that show once individuals learn to identify novel speech stimuli their FFRs are enhanced relative to more novice listening states (Cheng et al., 2021; Reetzke et al., 2018).

A possible mechanism that would enable FFRs to show real-time category representations is attention/behaviorally-dependent control of the corticofugal pathway. Attention heavily modulates responses from cortical structures (Bidelman and Walker, 2017; Harris et al., 2012;

Hillyard et al., 1973; Zhang et al., 2014). It is perhaps expected then that categorical representations in the (cortical) event-related potentials (ERPs) are only observed under states of attentional load and active speech labeling tasks (Alho et al., 2016; Bidelman and Walker, 2017; Carter, 2018). Literature on attentional effects in human brainstem responses is mixed with some suggesting attentional enhancement of FFRs (Galbraith et al., 1998; Hartmann and Weisz, 2019; Price and Bidelman, 2021) while others finding little to no effect of attention on the FFR (Aiken and Picton, 2008; Dunlop et al., 1965; Galbraith and Kane, 1993; Varghese et al., 2015). If attention does influence the brainstem FFR, then actively categorizing speech during behavioral tasks should yield measurable changes in the response. Moreover, stimulus order effects in the FFR would provide new evidence that subcortical speech representations are not only influenced by local stimulus history but are indeed tuned by nonlinear perceptual dynamics as observed at a cortical level (Carter et al., 2022).

The current study aimed to evaluate (1) if speech representations, as indexed by brainstem FFRs, show evidence of categorical representation or are strictly sensory-acoustic depictions of the speech signal; (2) whether attention and the process of categorization actively modulate speech-FFRs; (3) the effects of nonlinear dynamics (i.e., perceptual warping) on brainstem representations for speech. To this end, we measured speech-FFRs while listeners performed a rapid phoneme identification task where tokens along an identical categorical continuum were presented in random vs. serial (forward or backward) order. This design allowed us to induce more/less perceptual warping to bias listeners' categorical hearing. Serial order warps the perceptual space and corresponding *cortical* acoustic-phonetic representations for speech (Carter et al., 2022). Here, we determined if *subcortical* FFRs similarly carry category-level information that also changes with listeners' ongoing speech percept. We measured F0 and F1 attributes from FFRs to quantify "voice pitch-" and "formant timbre-" related coding in neural responses. We first confirmed our novel paradigm shifted individual's perceptual category boundary measured behaviorally and thus successfully warped (biased) listeners' percept. If brainstem speech coding is sensitive to categorization, we hypothesized FFRs would show larger amplitudes in sequential vs. random presentation orders due to listeners better predicting the phonetic category of the token in the former condition, resulting in enhanced phase locking and thus stronger FFRs. We anticipated these effects largely at the F1 frequency given this component was the primary cue for category labeling in our stimuli. However, pitch (F0) and timbre (F1) can interact in the speech-FFR, resulting in a modulatory effect on F0 dependent on the speech phoneme's identity (Krishnan et al., 2011). Thus, while we predicted it would remain invariant across tokens, it was conceivable that F0 might also vary in a category-dependent manner. We also anticipated relationships between neural and behavioral measures if the FFR is indeed modulated by listeners' ongoing categorical percept. Our data reveal that category-level features of speech are actively coded in FFRs, suggesting the acoustic-phonetic mapping of speech occurs more peripherally in the auditory system than previously thought.

2. Materials & methods

2.1. Participants

The sample included N = 16 young participants (24.2 ± 4.4 years; 5 females) averaging 16.9 ± 3.2 years of education; n = 9 of these listeners also participated in Carter et al. (2022). All were native American English speakers, had normal hearing (air conduction thresholds ≤20 dB HL; 250–8000 Hz), minimal musical training (≤3 years; average = 0.9 ± 1.2 years), and were mostly right-handed (mean = 78% ± 29% laterality) (Oldfield, 1971). Each gave written informed consent in compliance with a protocol approved by the University of Memphis IRB.

2.2. Stimuli & task

We used a synthesized 7-token vowel continuum from /u/ to /ɑ/. Each 100 ms token had a fundamental frequency of 150 Hz to minimize cortical contributions to the FFR which are restricted to low (< 100-120 Hz) stimulus frequencies (Bidelman, 2018b; Brugge et al., 2009; Guo et al., 2021). While there is still controversy in the literature regarding the FFR's generators (Bidelman, 2018b; Bidelman and Momtaz, 2021; Coffey et al., 2019; Gorina-Careta et al., 2021; López-Caballero et al., 2020), we have shown empirically that these 150 Hz vowel stimuli yield no identifiable "cortical FFR" sources and originate from brainstem sources (Price and Bidelman, 2021). That cortical contributions to the FFR are silenced for frequencies >150 Hz is further supported by other converging near- and far-field electrophysiological data (Bidelman, 2018b; Guo et al., 2021). Adjacent tokens were separated by equidistant steps in first formant (F1) frequency spanning from 430 (/u/) to 730 Hz (/ɑ/). We selected vowels over consonant-vowel (CV) syllables because our prior work showed vowels were more prone to nonlinear perceptual effects than stop consonants (Carter et al., 2022). We delivered stimuli binaurally through insert earphones (ER-2) at 80 dB SPL using rarefaction polarity with shielding to prevent stimulus electromagnetic artifact from contaminating neural responses (Campbell et al., 2012; Price and Bidelman, 2021). Sound presentation was controlled by MATLAB coupled to a TDT RZ6 signal processor (Tucker-Davis Technologies, Alachua, FL).

FFRs are sub-microvolt signals that typically require ~1000 trials to fully stabilize for the speech stimuli used here (Bidelman, 2018a). To use our categorization paradigm while simultaneously recording FFRs, we employed a modified version of the clustered interstimulus interval (ISI) presentation paradigm as described in Bidelman (2015c). This grouped stimuli in blocks containing rapid bursts of the same token (20 repetitions; ISI = 10 ms) within a short train. This fast rate also limited the likelihood of cortical contributions to the response (Chandrasekaran and Kraus, 2010). After each train, the participant selected the phoneme they perceived in the group with a binary keyboard response ("u" or "a"), after which the ISI was slowed (ISI = 400 ms) before the next grouping. The clustered ISI sequence was then repeated to achieve the appropriate trial counts to detect the FFR (x1000 presentations per token per condition) and sufficient behavioral responses (x50 per token) (Fig. 1).

There were three active conditions based on how tokens were sequenced: (1) random presentation, and two sequential orderings presented serially between continuum endpoints and F1 frequencies (2) forward /u/ to /ɑ/ (430 Hz to 730 Hz), and (3) backward /ɑ/ to /u/ (730 to 430 Hz). Forward and backward directions through the continuum were expected to produce perceptual warpings (i.e., hysteresis) (Tuller et al., 1994). An additional passive condition in which the stimuli were presented in a random order while the listeners watched a captioned film of their choice (but ignored the vowel stimuli) was used to test for attention effects on the FFR. The conditions were pseudo-randomly assigned using a Latin Square counterbalance (Bradley, 1958). In a subset of listeners (n=5), we measured the noise floor of our FFR recording setup to further rule out electromagnetic contamination of the neurophysiological recordings (see Fig. 6). This used an identical setup to the passive block only with the earphone removed from listeners' ear canal thereby recording only "neural noise" (Price and Bidelman, 2021).

2.3. EEG recording procedures

Neuroelectric activity was recorded between Ag/AgCl electrodes placed on the high forehead scalp (~Fz) referenced to linked mastoids (M1/M2) (with a mid-forehead electrode as ground), as is common for recording brainstem FFRs (Bidelman et al., 2013; Billings et al., 2019; Gockel et al., 2013; Price and Bidelman, 2021; Shukla and Bidelman, 2021). While other FFR montages are possible (Skoe and Kraus, 2010; Tichko and Skoe, 2017), this configuration is optimal for picking up the vertically oriented dipole(s) in the brainstem which pro-

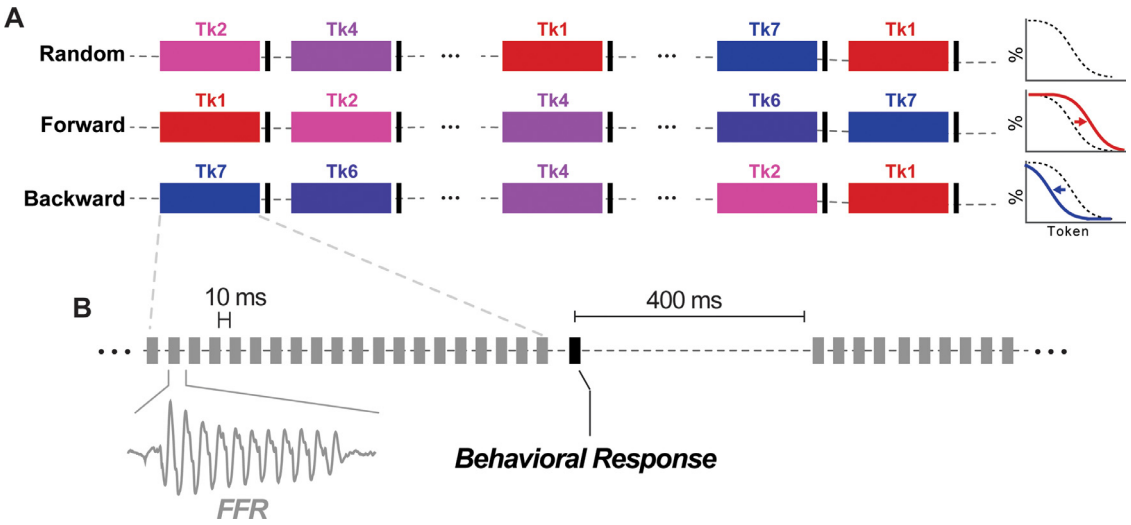

Fig. 1. Schematic of the stimulus clustering paradigm for recording FFRs during active behavioral tasks [modified from Bidelman (2015c)]. (A) Schematic of the stimulus presentation ordering and expected changes to speech identifications functions. Random vs. serial orderings (forward, backward) were achieved by presenting tokens from the speech continuum in clustered presentations (colored blocks; magnified in panel B) and varying the order between successive token cluster blocks. Continuum tokens are identified here as a red=/u/ to blue=/ɑ/ color gradient). Serial vs. random order produces a shift in listeners categorical boundary locations due to perceptual warping. (B) Zoom of each token cluster shown in panel B. Speech tokens were presented rapidly in blocks of twenty (10 ms ISI) to evoke the FFR. At the end of the block, stimuli were paused, and the listener categorized the sound as /u/ or /ɑ/. Following the behavioral response, a 400 ms pause occurred and the next block was presented. The clustered ISI sequence was then repeated to achieve the appropriate token counts for the FFR (x1000 presentations per token per condition) and sufficient behavioral responses (x50 per token).

duce a scalp topography to the FFR that is maximal near the Fz electrode (Bidelman, 2015b; Bidelman and Momtaz, 2021). Interelectrode impedances were kept ≤3 kΩ, except for three participants who had impedances ≤6 kΩ. EEGs were digitized at 10 kHz. We used BESA Research 7.0 (BESA, GmbH) to preprocess the EEG data. Responses were epoched (-5 – 105 window) and band-passed filtered (130–2000 Hz). This passband effectively attenuates cortical activity of the EEG while maintaining the high spectral resolution of the speech-FFR including the voice F0 and its harmonics captured in the response (Bidelman et al., 2013; Musacchia et al., 2008). We then utilized BESA's automated artifact rejection scan to automatically reject artifactual trial epochs exceeding a ±75 μV threshold criterion ($peak_{min}$-to-$peak_{max}$ voltage difference). This threshold was then manually adjusted in some subjects to retain at least 95% of the trials. Importantly, average trial counts across the three active stimulus conditions were within <10 trials of one another (random = 981.3 ± 16.8, forward = 987.7 ± 8.9, backward = 989.8 ± 8.4 trials). Additionally, formal measurements of response "neural noise" (i.e., pre-stimulus baseline RMS amplitude; Bidelman et al., 2014a; Krizman et al., 2021a) confirmed noise levels were similar across the three active conditions [$F_{2,320}$ = 0.0448, p = 0.95]. This rules out the possibility that differences in FFRs across conditions are due to trivial differences in recording quality. Clean trials were then averaged to derive FFRs for each vowel, stimulus direction, and participant.

2.4. Behavioral data analysis

2.4.1. Psychometric function analyses

Identification scores were fit with the sigmoid $P = 1/[1 + e^{-\beta 1(x-\beta 0)}]$, where P is the proportion of trials identified as a given vowel, x is the step number along the continuum, and $\beta 0$ and $\beta 1$ are the location and slope of the logistic fit estimated using non-linear least-squares regression (Bidelman and Walker, 2019; Bidelman et al., 2014c). Leftward/rightward shifts in $\beta 0$ location for the sequential vs. random stimulus orderings would reveal changes in the perceptual boundary characteristic of perceptual nonlinearity (Tuller et al., 1994). RTs greater than 2500 ms were considered outliers (e.g., attention lapses) and were excluded from analysis (reject trials: 208; 1.23% across all conditions/subjects/tokens) (Bidelman et al., 2013; Bidelman and Walker, 2019). We included RTs ≤250 ms, as we expected the task to induce faster RTs given a quasi-priming (anticipation) effect of the stimulus sequencing where the listener might decide on their percept during the ongoing stimulus train.

2.5. Electrophysiological data analysis

2.5.1. FFR analysis

FFR analyses were conducted using automated routines coded in MATLAB. We computed the Fast Fourier Transform (FFT) of each FFR to assess spectral content in each waveform. We then measured the F0 and F1 of the spectra as the maximal FFT amplitude in a window ±50 Hz around the nominal stimulus F0 and F1 frequencies. This was done to accommodate any variability between the neural response and the acoustic signal, allowing for potential warping of the FFR F0/F1 frequency due to category coding. As voice pitch (F0) was identical across our stimuli, we expected FFR F0 to remain invariant across tokens and sequence orders. In contrast, we expected differences in FFR F1 amplitudes where the stimuli are systematically changed to create the categorical continuum. We compared the FFT amplitudes of F0 and F1 of the same stimulus in different presentation conditions. Although not indicative of categorization, we also expected differences in F1 frequency across tokens since the FFR closely tracks changes in stimulus acoustics and we changed F1 frequency by the stimulus design (see Fig. 2d; Bidelman et al. (2013)).

2.5.2. Neural adaptation

As a control analysis to determine if neural adaptation occurred given the repetitive nature of our stimulus trains, we compared the F0 amplitude of the first and the last response of each train, across all tokens and conditions (2 × 50 × 7 × 4 = 2800 trials). A reduction in response ampltitude would indicate the fast repetition of our speech stimuli caused adaptation of neural responses (Pérez-González and Malmierca, 2014). Adaptation might inadvertently account for differential amplitude changes with stimulus presentation order and confound

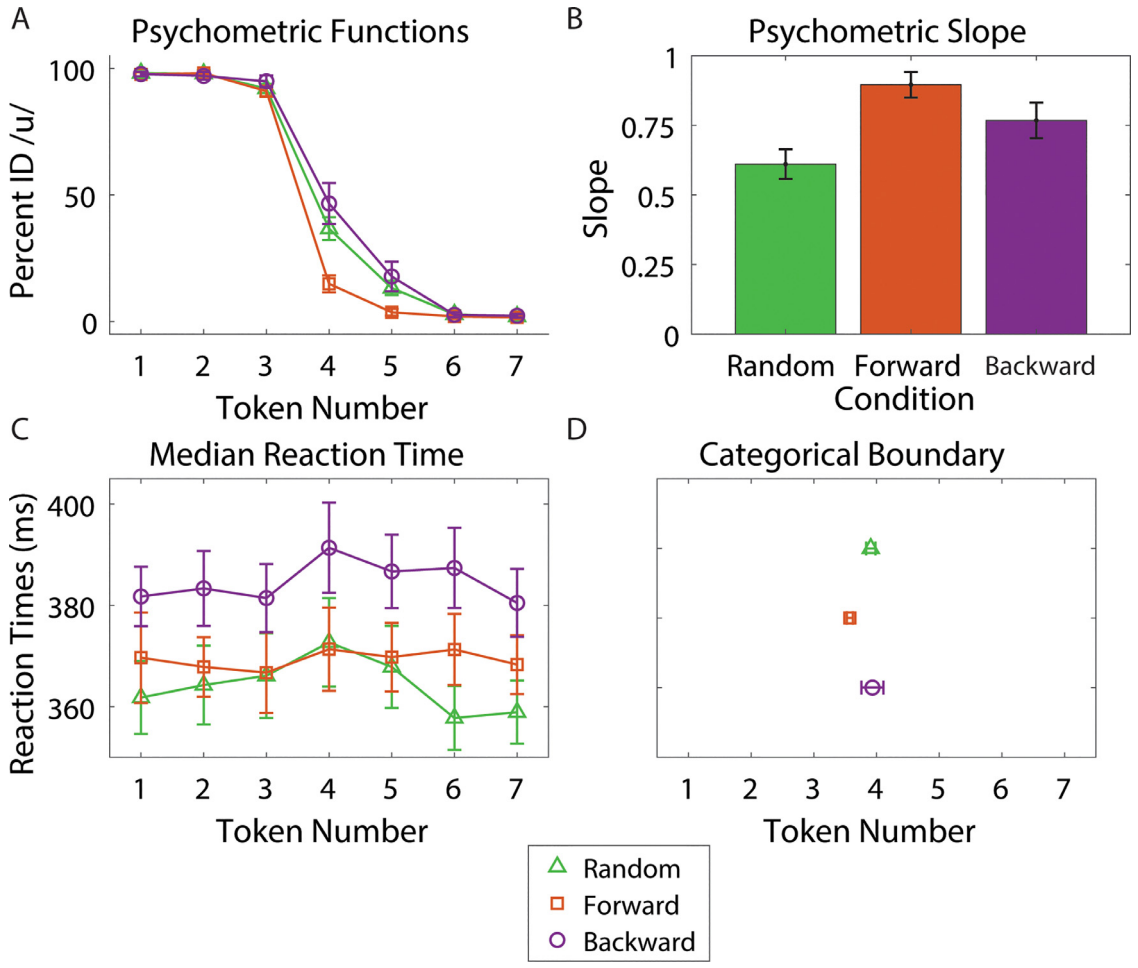

Fig. 2. Group level behavioral categorization. (A) Perceptual psychometric functions for phoneme identification when continuum tokens are presented in random vs. serial (forward: /u/→/ɑ/ vs. backward: /ɑ/→/u/) order. (B) Psychometric function slope was steeper for serial (forward and backward) compared to random presentation order. (C) Reaction times for speech identification. Backward presentation led to slower RTs than random and forward presentations. Additionally, there was no token difference in RTs. (D) Boundary location did not vary at the group level (cf. individual differences; Fig. 3). Errorbars = ±1 s.e.m.

our interpretations of hysteresis and categorical representations in the FFR.

2.5.3. Response-to-response correlations

To determine if stimulus ordering and thus perceptual warping biased listeners' speech-FFRs we measured response-to-response correlations between FFRs to the ambiguous token (Tk4) and the two endpoints (Tk1, Tk7) (cf. Yellamsetty and Bidelman, 2019). For each listener, we cross-correlated their time waveform to Tk4 for each serial order (forward, backward) with their time waveforms to both prototypical endpoints (Tk1 and Tk7 in the random condition). Waveforms were allowed to shift up to ±10 ms relative to one another to account for differences in delays (Galbraith and Brown, 1990). This resulted in four correlation coefficients per listener, reflecting the degree to which the FFR to the otherwise identical speech sound (Tk4) mirrored each of the two categories (i.e., Tk1 or Tk7). We reasoned that if the ambiguous token is more like one of the prototypical tokens than the other as a function of direction, it would indicate that the encoding of the signal was modulated by the perceptual warping induced by recent stimulus history (Yellamsetty and Bidelman, 2019). We additionally ran correlations between Tk1/7 and $Tk4_{FOR/BACK}$ as further validations and control analyses.

2.5.4. Neural decoding of speech categories from single-trial FFRs

We used Gaussian kernel, linear support vector machine (SVM) classifiers to determine if the categorical identity of speech stimuli could be decoded via FFRs (e.g., Lai et al. 2022; Xie et al. 2019; Yi et al. 2017). We used log-transformed F0/F1 amplitudes as the input features for SVM decoding. Amplitude measures were selected, as opposed to F0/F1 frequencies, as the latter would be trivial in separating FFRs since they closely follow the low-frequency pitch and formant cues of speech (Bidelman et al., 2013). Two binary classifications were considered: models assessing the classification of (i) FFRs elicited by the two category prototypes (i.e., Tk1=/u/ vs. Tk7=/ɑ/) and (ii) FFRs elicited by the otherwise ambiguous Tk4 in forward vs. backward serial directions. The first model was used largely as a control analysis since we expected robust separability of FFRs to differing vowel tokens and thus near ceiling classifier performance. In these analyses, for example, the classifier attempted to predict the FFR response on a given trial ($F0amp_n$) as being evoked by either an /u/ or /ɑ/ stimulus. Of more interest was the second model, which tested whether FFRs to a category-ambiguous speech sounds were warped based on listeners' trial-to-trial phonetic hearing.

For each classifier, we extracted F0 and F1 amplitudes from *single-trial* FFR spectra, resulting in N=27000 observations per condition. We randomly split the data into training (80%) and test (20%) sets (Mahmud et al., 2021). We then trained an SVM via the *fitckernel* function in MATLAB using the default box constraint (*C=1*) and regularization ($\lambda = 1/n$ = 4.63e-05) tuning parameters, where C = $1/(\lambda n)$. This algorithm maps data from a low- to high-dimensional space, then fits a linear model in the high-dimensional space by minimizing the regularized objective function. We used 5-fold cross validation to prevent

model overfitting (Mahmud et al., 2021). In this procedure, the dataset is partitioned into k equal-sized subsamples (folds) containing $N = 21600$ (80% training) and $N = 5400$ (20% testing) observations. For each fold, the SVM learned the support vectors from the training data that optimally segregated the FFR attributes (i.e., F0 and F1 amplitudes) based on the class labels (e.g., Tk1 vs. Tk7 or $Tk4_{for}$ vs. $Tk4_{back}$). This was repeated for each fold. The final classifier performance represented the mean decoding accuracy (i.e., % of matches between predicted and true class labels) averaged across folds. Classification performance was then assessed using conventional classifier metrics [i.e., accuracy, d-prime, receiver operating characteristic curve (ROC), confusion matrices].

2.6. Statistical analysis

We used one-way mixed model ANOVAs (PROC GLIMMIX, SAS® 9.4; SAS Institute, Inc.) to analyze the psychometric data, with a fixed effect for presentation condition (3 levels: random, forward, and backward), and a random effect for subjects. RTs and FFR data (i.e., F0 and F1 frequency and amplitudes) were analyzed using two-way, mixed model ANOVAs (subjects = random factor) with fixed effects of condition (3 levels: random, forward, backward; 4th level for FFR: passive) and token (7 levels). We normalized the heavily bimodal distribution of the correlation data by taking the absolute value of the difference of the individual's correlation value and the mean of all correlations [i.e., abs(X - mean(X)].

We used orthogonal quadratic trend contrasts on F0 and F1 amplitude measures to test for the characteristic U-shape pattern inherent to categorical responses (Pisoni, 1973). These *a priori* contrasts (coefficients = 5, 0, -3, -4, -3, 0, 5) assessed whether FFR amplitudes to token prototypes were larger (or smaller) than ambiguous tokens near the continuum's midpoint (Carter and Bidelman, 2021; Mankel et al., 2020) and therefore differentiated speech sounds with strong vs. weak category percepts. We anticipated that if category-level information is encoded in brainstem responses, a similar quadratic trend would arise.

We conducted general linear mixed effects (GLME) regression models (*fitglme* in MATLAB) to assess whether a linear combination of the neural measures (i.e., F0/F1 frequencies and amplitudes) predicted behavior [e.g., behav ~ FFR_{F0amp}+ FFR_{F0freq} + FFR_{F1amp}+ FFR_{F1freq} + (1|sub)]. Subjects served as a random factor in these models. Separate GLMEs were run for each behavioral metric (i.e., slope; boundary; RTs). Responses across the three orders were pooled for data reduction.

3. Results

3.1. Behavioral data

Listeners perceived the vowels categorically in all presentation orderings as seen in Fig. 2. Slopes varied with presentation order ($F_{2,30} = 11.21, p = 0.0002$). The random condition was significantly shallower than both the forward ($p = 0.0001$) and backward ($p = 0.0367$) conditions. The location of the categorical boundary only showed marginal shifts with presentation order at the *group level* ($F_{2,30} = 3.14$, $p = 0.0576$). These findings are consistent with notions that categorical speech percepts are stronger when stimuli are presented in a sequential compared to random order (Carter et al., 2022).

RTs also varied with presentation order ($F_{2,312} = 18.27, p < 0.0001$). Categorical decisions were slower for backward versus forward ($p < 0.0001$) and random ($p < 0.0001$) presentations. This finding indicates that the backward condition slowed processing speed in categorization. However, there was no difference in the RTs between ambiguous and prototypical tokens ($F_{6,300} = 1.38, p = 0.22$). This suggests that under our clustered stimulus paradigm, listeners may have decided the category while the stimulus train was still ongoing. Additionally, no interaction occurred between presentation order and token ($F_{12,300} = 0.38$, $p = 0.97$). While the group level categorical boundary was largely stagnate, individual-level data showed stark differences as a function of presentation order (Fig. 3).

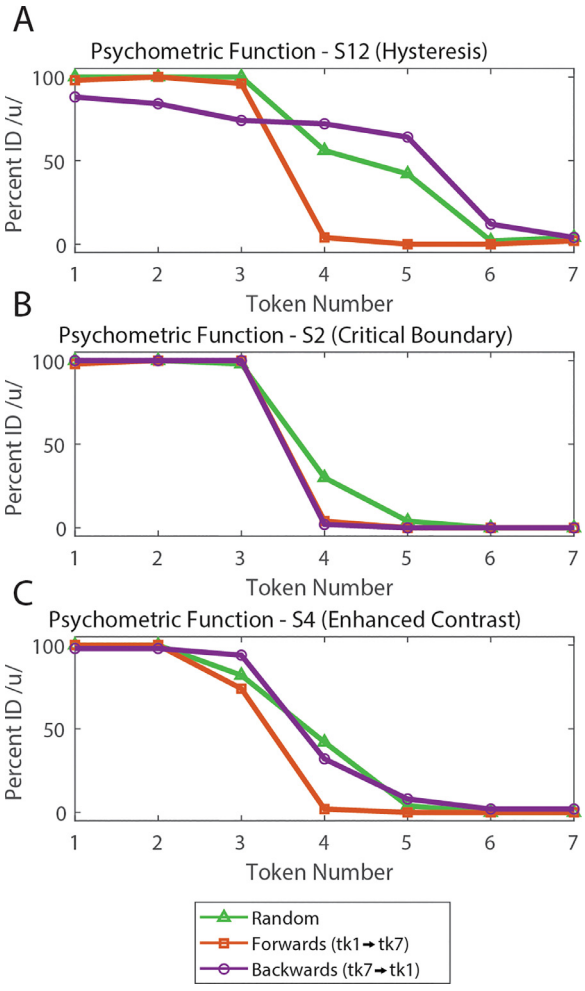

Fig. 3. Individual level psychometric functions. Representative subjects (n=3) who showed (A) hysteresis (B) critical boundary and (C) enhanced contrast perceptual response patterns.

3.2. Electrophysiological data

Figs. 4 and 5 show time-domain FFR waveforms for select conditions and tokens. These waveforms were analyzed in the frequency domain to determine differences in F0 and F1 frequency and amplitude.

Fig. 6 shows FFR spectra in response to Tk1 across stimulus orderings (random, forward, backward) and attention conditions. Fig. 7 shows F0 and F1 measures more clearly. We found that F0 amplitude differed as a function of condition ($F_{3,405} = 7.78$, $p < 0.0001$) and token ($F_{6,405} = 5.76$, $p < 0.0001$). Post-hoc testing revealed the passive F0 amplitudes were smaller than all three active listening conditions (backward, $p = 0.0005$; forward, $p = 0.0047$; random, $p = 0.0001$). The token effect was attributed to smaller F0 amplitudes in response to /u/ tokens (Tks 1-3) compared to /ɑ/ tokens (Tks 5-7) ($p < 0.0001$). Conversely, F1 amplitude did not differ as a function of condition ($F_{3,405} = 0.31$, $p = 0.8189$), but did as a function of token ($F_{6,405} = 131.79$, $p < 0.0001$). These results indicate (expectedly) the FFR is sensitive to the acoustic properties of speech across the stimulus continuum. More critically, they indicate subcortical speech representations are enhanced with active attention. The F0 frequency also shifted due to stimulus orderings ($F_{3,405} = 3.95$, $p = 0.0086$; order x token interaction: $F_{18,405} = 1.82$,

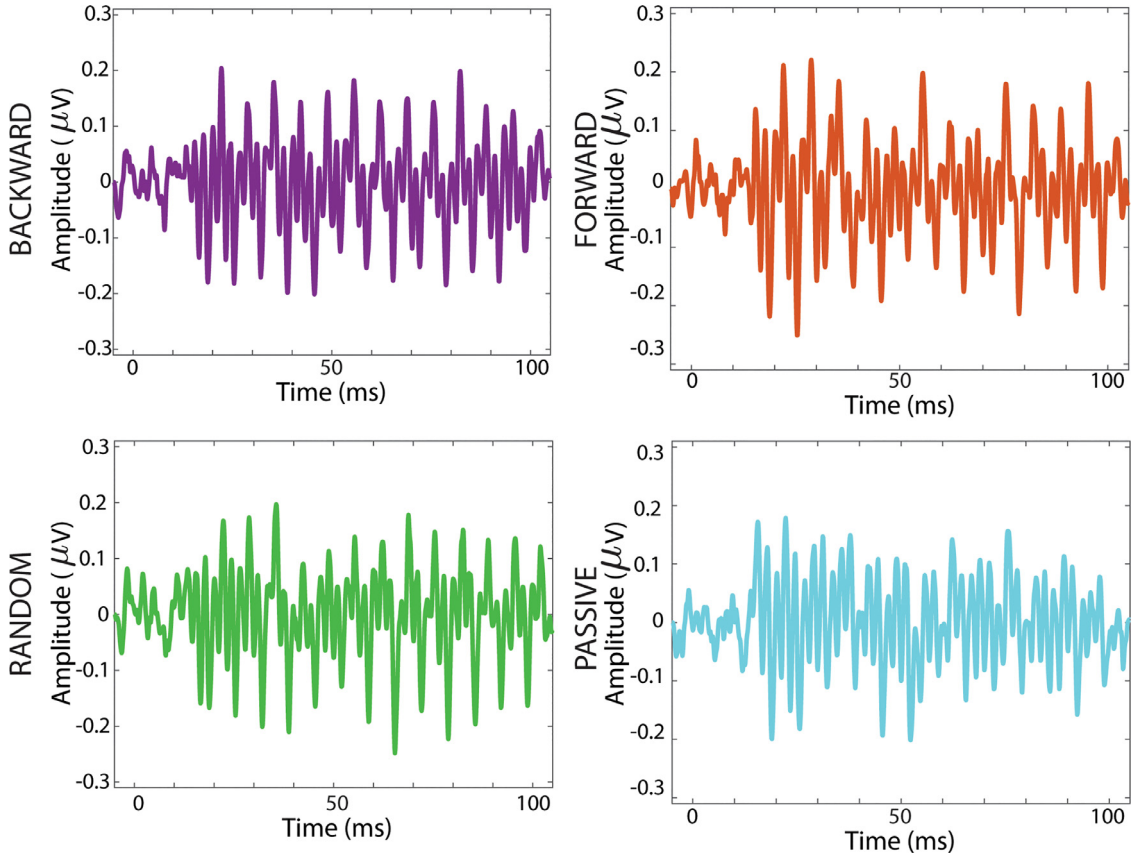

Fig. 4. FFR time domain waveforms (Tk1) contrasting stimulus presentation orders and attentional state (i.e., active vs. passive listening).

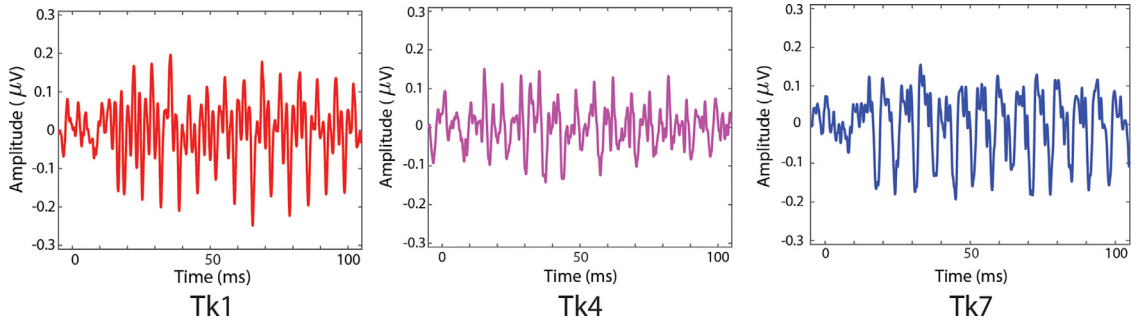

Fig. 5. FFR time domain waveforms comparing responses to mid- and endpoint tokens (Tk1, Tk4, Tk7) for the random condition. Note the larger FFR amplitudes for the endpoint vs. midpoint token indicative of a neural enchancment to category prototypes.

$p = 0.02$). However, this effect was due to a small but measurable ~5 Hz increase in FFR F0 for the passive relative to other conditions. In contrast, F1 frequency was not significantly affected by presentation order ($F_{3,405} = 0.04$, $p = 0.99$).

We found FFRs across the categorical continuum displayed a quadratic trend for both F0 ($F_{6,405} = 5.76$, $p < 0.0001$) and F1 ($F_{6,405} = 131.79$, $p < 0.0001$) measures. Quadratic trends showed a U-shape for the F0 amplitudes in the backward ($p = 0.0316$) and forward ($p = 0.0149$) conditions, but not for the random ($p = 0.1651$) or passive ($p = 0.5883$) conditions. In terms of effect size (Cohen's-*d*), serial orderings produced a much stronger U-shape in the F0 data than either the random or passive conditions ($d_{for} = 5.79$, $d_{back} = 5.10$, $d_{rand} = 3.29$, $d_{passive} = 1.28$). Said differently, the serial orderings produced a pattern of responses consistent with perceptual warping that was ~1.65x SDs larger than that of the random condition and upwards of ~4.5 SDs larger relative to passive condition. These results suggest that, despite identical F0s in the stimuli, the FFR showed categorical coding of F0 only in sequential presentation orders (which elicited perceptual warping). For F1, quadratic trends were highly significant ($p < 0.0001$) for the F1 amplitudes in all conditions ($d_{for} = 13.2$, $d_{back} = 15.6$, $d_{rand} = 14.2$, $d_{passive} = 15.1$). These results suggest that the FFR F1 also showed categorical coding regardless of attention or presentation order. However, in contrast to F0, visual inspection of the F1 trends appeared less U-shaped and largely dominated by stronger responses at the /u/ (low-) vs. /ɑ/ (high-frequency) end of the continuum (Supplemental Fig. S3), which could be explained by the roll off in phase-locking. Critically however, neither the F0 nor F1 quadratic patterns were observed in the physical acoustic stimuli (Supplemental Fig. S4) nor FFRs simulated from a well-established model of the auditory nerve that captures important cochlear processing including spectral decomposition and compressive nonlinearities (Bidelman, 2014; Zilany et al., 2014) (Supplemental Fig. S5). In general, model FFRs across tokens more closely mirrored the pattern observed in the stimulus acoustics rather than actual FFRs. These findings suggest, at least qualitatively, that the category coding effects

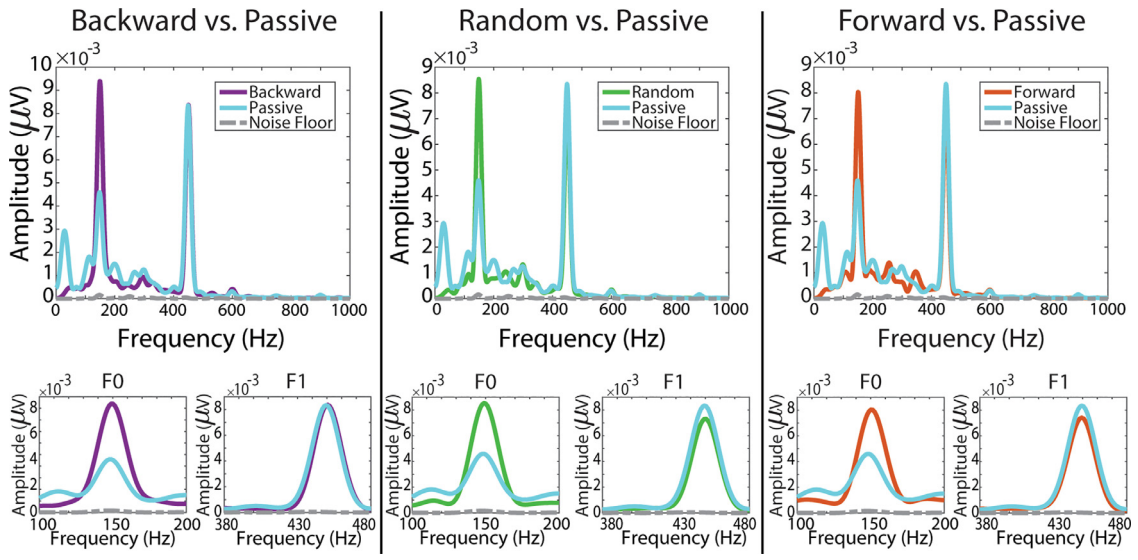

Fig. 6. FFR spectra (Tk1) in the backward, random, and forward conditions vs. the passive condition. Insets show F0 and F1 analysis windows.

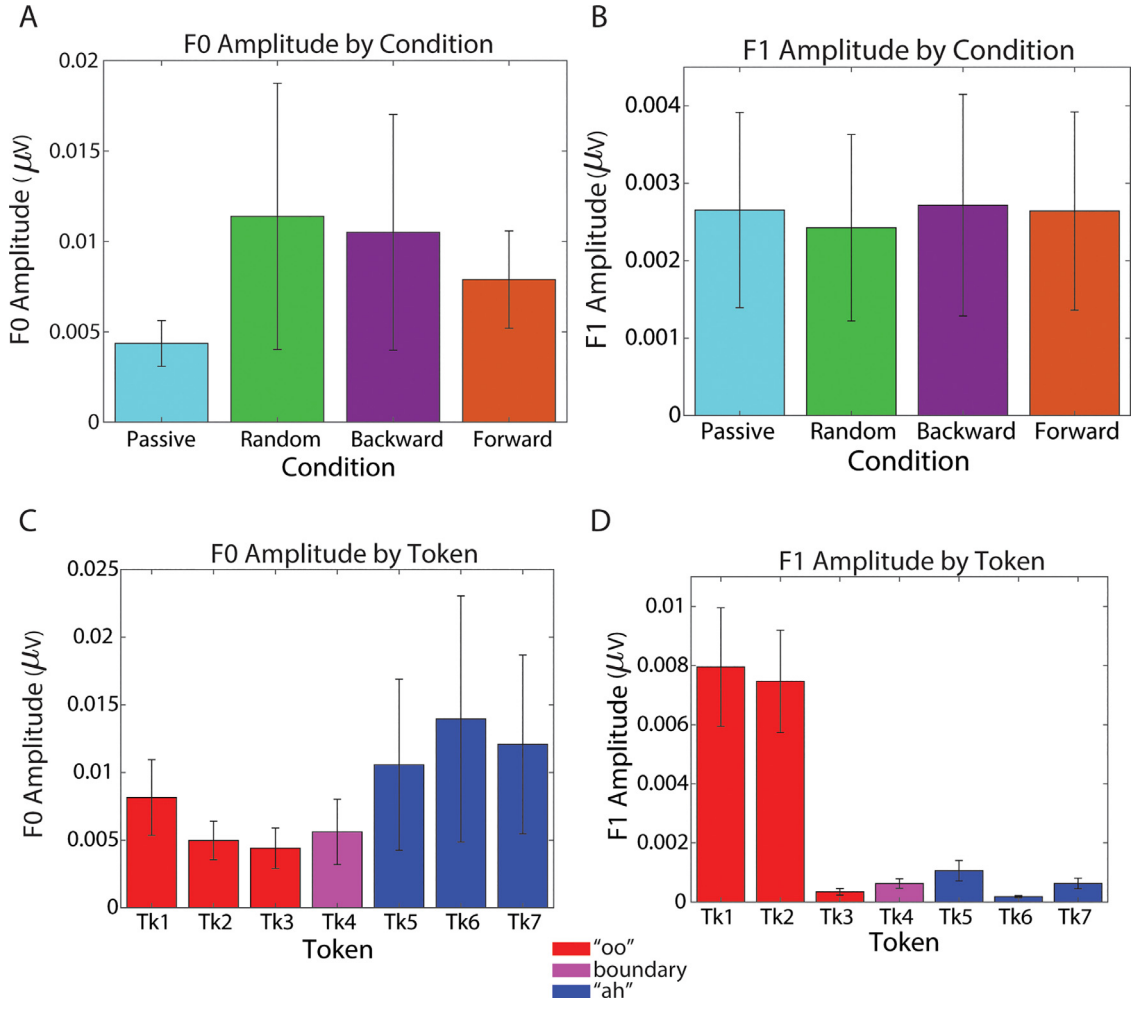

Fig. 7. FFR F0 and F1 measures as function of stimulus order and token. (A) The F0 amplitude of active conditions (i.e., random, forward, and backward) were greater than the F0 amplitude in the passive condition. (B) F1 amplitudes were not affected by stimulus direction nor attention. (C) F0 amplitudes (pooling orders) showed a U-shape pattern suggesting categorical coding across speech tokens (Pisoni, 1973). (D) F1 amplitudes (pooling orders) were significantly larger for /u/ vs /ɑ/ ends of the continuum. Errorbars = ±1 s.e.m.

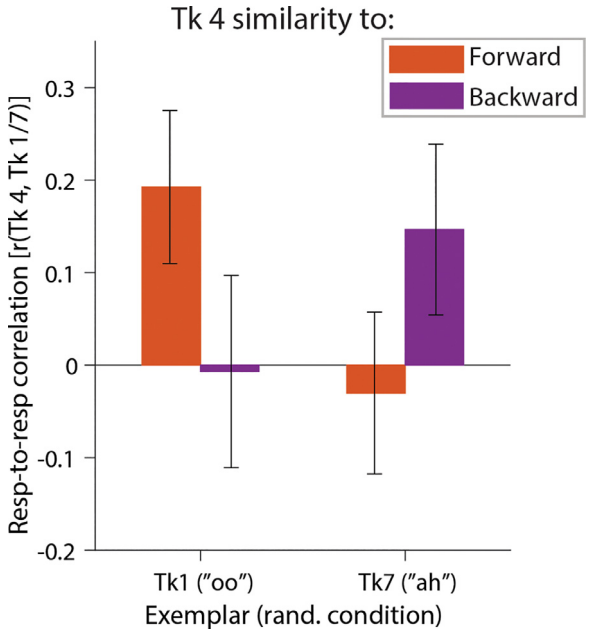

Fig. 8. FFRs show category-specific coding. Comparison of response-to-response correlations between FFRs to the ambiguous speech token (Tk4) presented in backward and forward conditions with responses to either prototypical vowel (Tk1/7). Higher correlation coefficients indicate a stronger similarity to that speech category (i.e., /u/ or /ɑ/). Errorbars = ±1 s.e.m.

observed in empirical FFRs are not due to stimulus acoustics or the output of cochlear transformations, *per se*, but instead reflect central, top-down modulations from listeners' perceptual state and attention.

We measured the F0 of the first and last tokens in each train to quantify possible neural adaptation to the rapid stimuli in our clustered presentation paradigm. We found no difference in F0 amplitude between the first and last response in each train ($F_{1,615}$ = 0.20, p = 0.6551; Supplemental Fig. S1). This confirms there was little to no adaptation of brainstem responses due to the rapid succession of auditory stimuli (Bidelman and Powers, 2018) and thus rules out the confound that serial order effects in the FFR data were due to mere neuronal fatigue.

Fig. 8 shows response-to-response correlations between Tk4 (ambiguous token) and Tk1/7 (prototypical tokens) FFRs as a function of presentation order. We found a main effect of presentation order ($F_{1,45}$ = 4.39, p = 0.0417) and a significant interaction between presentation order and token ($F_{1,45}$ = 4.92, p = 0.0317). The interaction suggests Tk4 FFRs showed stronger similarity to Tk1 in the forward direction but stronger correspondence to Tk7 in the backward direction. By token, the direction contrast was stronger at Tk1 (p = 0.0038) than Tk 7 (p = 0.93). These findings suggest that the FFR to an otherwise identical (and categorically ambiguous) speech token was modulated by perceptual state. That is, FFR neural representations were warped toward the direction of the vowel prototype under each stimulus context (i.e., mirroring Tk1 for forward stimulus ordering and Tk7 for backward stimulus ordering).

In attempts to isolate whether F0 or F1 drove these effects, we bandpass filtered FFRs around each component (F0: 130-160 Hz; F1: 400-750 Hz) and recomputed the response-to-response correlations. The order x token interaction was observed for F0 ($F_{1,45}$ = 3.95, p = 0.052) but not F1 ($F_{1,45}$ = 2.32, p = 0.13). While marginal, these findings imply that F1 likely did not contribute to the perceptual coding observed in Fig. 8 and was therefore dominated by the F0 component.

To ensure these inter-FFR correlations were not trivially common across all tokens, we also conducted response-to-response correlations for Tk1/7, which was not significant (r_{mean} = 0.07, p = 0.4985), and $Tk4_{FOR/BACK}$ which was significant (r_{mean} = 0.32, p < 0.0001). These findings indicate that responses to acoustically (and categorically distinct) tokens were not correlated with one another and that responses to the same token carrying different perceptual encodings still retains the general acoustic encoding of the signal.

Neural classifier performance is shown in Fig. 9. Single-trial decoding on FFR amplitudes was expectedly robust for classifying the phoneme endpoints of the continuum (i.e., Tk1 vs. Tk7). At the group level, cross-validated accuracy was 86% AUC (d-prime=1.49), resulting in few confusions between true and predicted token labels and thus highly discriminable responses (X^2 = 8468, p < 0.0001). Individual vowel decoding from FFRs was equally good and well above chance levels (one-sample t-test against 50%: t_{14} = 27.62, p < 0.0001). These control decoding analyses indicate that phonetic properties of vowel prototypes (i.e., /u/ vs. /ɑ/) were easily distinguished via spectral amplitude features carried in FFRs.

Having confirmed FFRs carry sufficient information on the category identity of speech signals, we next asked whether responses to category ambiguous speech (Tk4) showed differential encoding depending on the direction of stimulus presentation. As confirmed behaviorally, forward vs. backward serial ordering of the stimulus continuum induced perceptual shifts that changed listeners' percept of otherwise identical speech sounds (see Fig. 2). Single-trial decoding was expectedly poorer for this more challenging classification problem given lower separability of the data. However, FFRs were surprisingly distinguishable based on stimulus order (i.e., $Tk4_{for}$ vs. $Tk4_{back}$). Group level classification accuracy was 62% AUC (d-prime=0.45) with more frequent vowel confusions—the SVM tended to predict more Tk4 responses as stemming from the backward condition, suggesting a slight bias in labeling. Still, the confusion pattern was highly discriminable (X^2 = 66.5, p < 0.0001) and individual decoding was well above chance (t_{14} = 9.26, p < 0.0001). These neural decoding results complement the response-to-response correlations and indicate that FFRs to otherwise identical speech stimuli are warped on a trial-by-trial basis according to the surrounding stimulus context. In subsequent analyses, we focus on the relevance of these dynamic neural effects to *behavioral* categorization of speech.

3.3. Brain-behavior relationships

We used GLME regression models to determine whether neural FFR measures predicted aspects of listeners' categorical perception. For $\beta 0$ (boundary location), the multivariate model indicated FFR measures predicted ~63% of the variance in behavior (adjusted R^2 = 0.63)[1]. Evaluating individual terms revealed a significant predictor in FFR F1 frequency on listener's categorical boundaries (t_{11} = 2.43, p = 0.03). For $\beta 1$ (psychometric slopes), the multivariate model predicted ~71% of the variance (adjusted R^2 = 0.71). Evaluating individual terms revealed a significant predictor in FFR F0 frequency on listener's psychometric slopes (t_{11} = 2.87, p = 0.015) (see Supplemental Tables S1 and S2). These results suggest that subcortical coding of different speech features predicts listeners' vowel categorization.

4. Discussion

We measured brainstem FFRs concurrent with behavioral responses to acoustic-phonetic continua presented in various stimulus orderings (sequential vs. random presentation) and attentional states (active vs. passive tasks). Our innovative stimulus task establishes a new paradigm to obtain FFRs and behavioral responses to speech concurrently. Using

[1] Adj-R^2 are adjusted for the number of fixed-effects in the model and are computed using the error (SSE) and regression (SSR) sum of squares. SSE is based on the conditional response of the linear mixed-effects model. Thus, MALTAB's `fitglme()` adj-R^2 takes into account both the fixed and random effects in the model. For details see, https://www.mathworks.com/help/stats/generalizedlinearmixedmodel-class.html

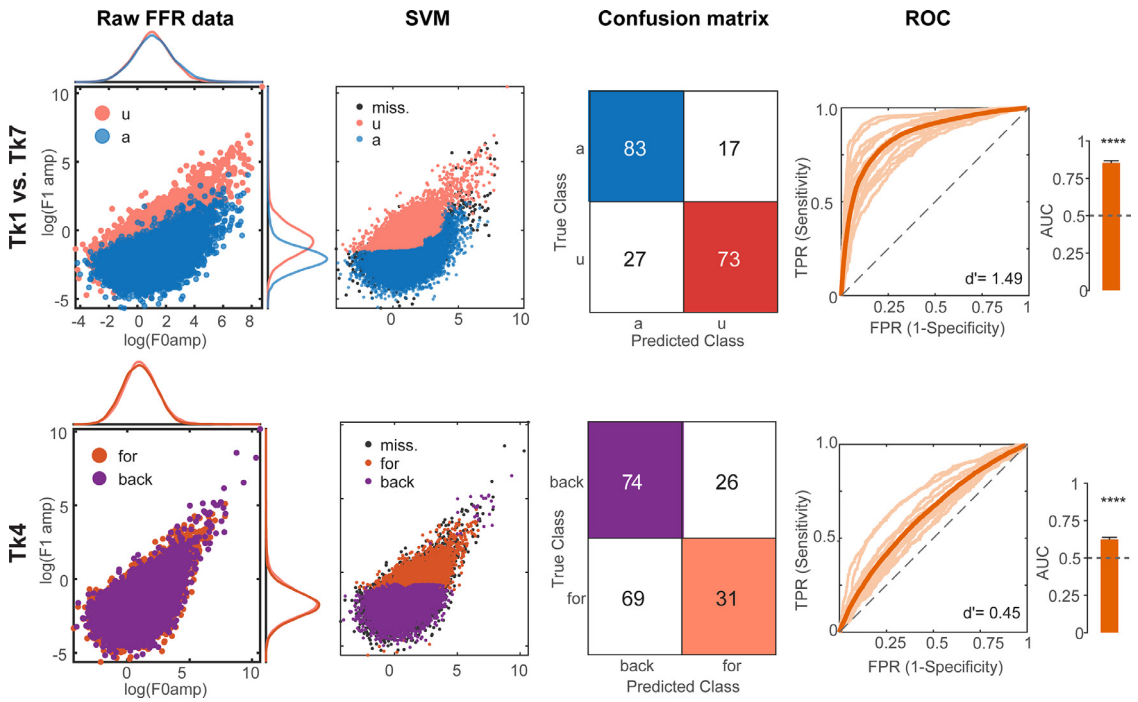

Fig. 9. Single trial FFR decoding reveals evidence of phonetic encoding and perceptual warping depending on stimulus presentation order. SVM classification of FFRs decoding (*top row*) category prototypes (Tk1=/ɑ/ vs. Tk7=/u/) and (*bottom row*) stimulus order (direction) for ambiguous Tk4 responses. *First column*, raw FFR F0 and F1 amplitudes extracted from N=27000 single trial FFRs. *Second column*, SVM output showing the decision boundary and posterior class labels predicted after SVM training. Black=misclassified observations. *Third column,* cross-validated confusion matrices show the proportion of predicted vs. true class labels. Higher values along the diagonal denote more successful separability of FFRs and better decoding performance. *Fourth column*, ROC curves. Thin lines= single subjects; thick line; grand average across participants; dotted line= chance performance. *Fifth column,* average ROC area under the curve (AUC) across listeners for classifier performance (cf. %-accuracy). (*Top row*) Prototypical categories (Tk1 vs. Tk 7) are easily distinguished via neural FFRs, resulting in few confusions and high classification accuracy. Note the large separability in F1 (but not F0) amplitude measures in the raw data. (*Bottom row*) Decoding FFRs to Tk4 presented in the forward vs. backward serial directions, which produce perceptual hysteresis. Decoding performance is well above chance even in light of the low separability of the data (i.e., all Tk4 trials) suggesting FFRs contain adequate information to code listeners' trial-by-trial speech percepts. TPR, true positive rate. FPR, false positive rate. AUC, area under the (ROC) curve. $^{****}p < 0.0001$; Errorbars = ±1 s.e.m.

this novel approach, we show that attention modulates the encoding of speech as early as the auditory midbrain and moreover, that FFRs encode speech categorically. Interestingly, while we anticipated changes would occur at *F1*, more salience effects were observed at *F0*. We suspect this may be due to the F0 component being a more dominant driver of the overall FFR waveform than the weaker amplitude F1 (e.g., Fig. 6) and thus more easily modulated by top-down influences. Indeed, prior studies show that corticofugal influences on the FFR are largely restricted to low-frequency portions of the speech signal (Lai et al., 2022; Price and Bidelman, 2021).

4.1. FFR responses obtained concurrent with active task

Most speech-FFR studies drawing putative links between auditory brainstem coding and aspects of speech perception have used passive listening tasks (Aiken and Picton, 2008; Bidelman et al., 2013; Skoe and Kraus, 2010; Slugocki et al., 2017). This has led to claims that FFRs reflect a perceptual correlate of behavior. However, in the absence of an active perceptual task in previous work, establishing this link is more difficult and speculative. Recent advancements in stimulus paradigms have shown that active, perceptual challenging tasks can induce modulations in the speech-FFR, revealing brainstem representations are subject to attentional gain modulation (Price and Bidelman, 2021). Through use of our innovative clustered stimulus paradigm, we further demonstrate a feasible method to obtain speech-FFRs simultaneous with an active behavioral speech listening task. Our data provide new and important evidence that speech-evoked brainstem responses, like their cortical ERP counterparts (Carter et al., 2022), are actively modulated by listeners trial-by-trial perception of the speech signal and its surrounding context. Consequently, we infer FFRs reflect more than mere sensory-acoustic representations, and instead carry true perceptual correlates of the speech signal.

Behaviorally, we found the slopes of listeners' psychometric functions were steeper in sequential vs. random presentation ordering. This agrees with previous findings (Carter et al., 2022) and suggests that sequential presentation solidifies categorization as individuals rapidly decide what category to assign the sound stimulus. Additionally, serial presentation of tokens in our paradigm likely strengthens the sensory (echoic) memory trace which would reinforce individuals' decision by the time they execute their behavioral response (Näätänen et al., 2007; Winkler et al., 1993).

Surprisingly, RTs were slower in backward vs. both the forward and random conditions. On the contrary, we would have expected the random condition to produce longer RTs than either sequential condition. RTs may have been slower in the backward condition due to a greater salience of rising than falling frequency stimuli (Carter et al., 2022; Luo et al., 2007; Schouten, 1985). Perhaps in our paradigm, listeners subtly slowed their identification to ensure they were selecting the correct sound, whereas in forward and random conditions, the change in F1 frequency was perceptually salient enough to keep RTs rapid. Additionally, RT patterns in conventional speech categorization tasks typically show an inverted U shape across the continuum, with RTs slowing around the categorical boundary compared to the prototypical tokens (Pisoni and Tash, 1974). We did not observe this in the current study. This may relate to listeners deciding on their percept early in the stimulus train, then selecting their response once the train ends. That is, RTs

might be locked more to the ending of the entire stimulus train than to the processing of the phoneme, *per se*. Despite the lack of token effect, RTs were however modulated by the overall ordering of speech, indicating that decision speeds can be facilitated by recent stimulus history (i.e., context).

4.2. Brainstem FFRs are modulated by attention

Strikingly, we found that speech-FFRs (F0 amplitudes) were much larger in active vs passive conditions, confirming that attention actively shapes neural encoding at the brainstem level. We had expected to also see changes in FFR F1 amplitudes as a function of presentation order, but this was not observed (see Fig. 7). Attention effects in the FFR thus seem localized to low-frequency components of the speech signal (Holmes et al., 2018). The effect of attention on any property of brainstem responses has been highly equivocal in previous work; some studies support (Galbraith et al., 1998; Hartmann and Weisz, 2019; Price and Bidelman, 2021) and others refute (Aiken and Picton, 2008; Dunlop et al., 1965; Galbraith and Kane, 1993; Varghese et al., 2015) attentional effects on FFRs.

Attentional modulation of FFRs observed here is presumably driven by corticofugal fibers that enhance brainstem activity selectively according to perceptually-relevant information in cortex. Animal studies have shown the corticofugal fibers shape subcortical function during short-term learning (Bajo et al., 2010; Suga, 2008). In humans, corticofugal mechanisms are thought to be particularly important in difficult speech-listening environments (Lai et al., 2022; Price and Bidelman, 2021). These effects could relate to the short-term memory modulation caused in nonlinear dynamical processing of speech, wherein the encoding of ambiguous speech tokens at lower levels are continuously shaped by higher cortical structures. Indeed, perceptual warping effects on primary auditory cortex responses are thought to arise from prefrontal memory areas (Carter et al., 2022). It is possible such perceptually-relevant biasing percolates back to even more peripheral auditory areas (i.e., brainstem) as suggested by the category tuning of FFRs observed here. In this regard, corticofugal fibers might carry category identity from cortex further down the system, rendering changes in speech representations at the brainstem level. Previous anatomical work has demonstrated cortico-collicular connections originating in the frontal lobes and terminating in the brainstem that contain GABAergic and glutamatergic neurons. These connections are thought to shape responses in inferior colliculus, a major source of the FFR, through complex excitatory and inhibitory interactions (Liu et al., 2023; Olthof et al., 2019). Consequently, the necessary circuitry is in place for higher-level brain regions (frontal lobe) to modulate early signal encoding in the FFR (e.g., Liu et al. 2023). Our data provide strong evidence of attentional modulation of subcortical responses, possibly originating in the distal frontal lobes. Though future studies are needed to confirm this hypothesis.

Our task requires listeners to perform online categorization judgments and continuously monitor the speech stimuli. Previous tasks evaluating brainstem-attention effects have used simple tasks (e.g., counting, detection, attention redirection, etc.) (Galbraith et al., 1998; Galbraith and Kane, 1993; Varghese et al., 2015) or oddball paradigms (Hartmann and Weisz, 2019; Price and Bidelman, 2021), which may allow listeners to periodically disengage from the task and fail to produce FFR-attention effects. We have recently shown that task disengagement has strong influences on cortical arousal which simultaneously causes fluctuations in speech FFR responses (Lai et al., 2022). Our task arguably requires more sustained attention which may account for the much larger (x2-3) brainstem attentional effects we find in the present study compared to previous reports (Price and Bidelman, 2021). Our findings agree with some studies demonstrating improvements in FFR encoding under active attention (Forte et al., 2017; Galbraith et al., 1998; Krizman et al., 2021b; Lai et al., 2022; Price and Bidelman, 2021); however, other studies have shown the opposite, with no attentional modulation of FFR encoding (cf. Galbraith and Kane, 1993; Hillyard and Picton, 1979; Varghese et al., 2015). Attentional effects in the FFR are stronger for low (93-109 Hz) vs. high (217-233 Hz) frequency stimuli, which suggests cortical components of the FFR—when present for low-frequency stimuli—might be more prone to attentional effects compared to those from subcortical sources (Holmes et al., 2018). Additionally, Hartmann and Weisz (2019) found that while attentional effects do occur in the FFR_{MEG} for low F0s, it is primarily responses from right primary auditory cortex that show attentional change. Our study differs from these two studies in our use of speech stimuli with higher frequencies (both F0 and F1) which generate FFRs from brainstem structures with little to no contribution from cortex (Bidelman, 2018b; Price and Bidelman, 2021).

4.3. Brainstem FFRs carry category-level information (perceptual correlates) of speech

Another novel finding revealed by our innovative task is that FFRs encode speech in a categorical fashion. Category representation in the FFR is unlikely to be local to the midbrain. Rather, we posit that corticofugal fibers modulate early sound encoding of the stimulus to fit the perception of the token (Suga, 2008; Suga et al., 2000). We have previously shown that at the level of cortex (Carter et al., 2022), activity in frontal brain regions influences the encoding and subsequent perception of category-ambiguous speech sounds (cf. Tk4) (Carter et al., 2022). Previous work has also demonstrated that changes in perception can drive enhancements of the FFR (Cheng et al., 2021), suggesting speech processing is influenced by predictions of the percept. Our response-to-response correlations and neural decoding results support perceptual encoding in the FFR. Brainstem responses to otherwise ambiguous speech tokens were biased towards a given prototype depending on the direction of presentation. These results were independent of neural adaptation ruling out explanations that our FFR warping effects were driven by the normal physiological byproducts of rapid auditory processing (i.e., refractory of neuronal firing). In contrast to our data, some animal studies have shown stimulus-specific adaptation in the auditory midbrain (Pérez-González et al., 2005). Nevertheless, the lack of adaptation in the present data suggests our FFRs are likely not cortical in nature, since cortical neurons would be expected to show significant response diminishment for the rapid stimulus rates used here. Together, this indicates the FFR is not merely a passive representation of the acoustic speech signal but is dynamically shaped by higher-order perceptual processes, and by surrounding stimulus context. Such active modulation of brainstem representations might help simplify speech decisions upon arrival to auditory cortex (Asilador and Llano, 2021; Lesicko and Geffen, 2022).

Importantly, our speech stimuli were designed with F0s well above cortical phase locking limits as reported in both humans and animal models (Brugge et al., 2009; Gnanateja et al., 2021; Guo et al., 2021; Joris et al., 2004; Wallace et al., 2000). This reduces the possibility that our FFR results are conflated by cortical contributions (Coffey et al., 2016). This, in addition to the lack of neural adaptation (characteristic of more peripheral auditory nuclei), strongly supports a brainstem locus of our findings.

It is useful to ask whether the pattern(s) observed in FFRs can be explained by acoustic factors of the stimuli or more peripheral signal transformations, e.g., due to cochlear nonlinearities. Given its more central generators, FFRs presumably reflect the combined output of peripheral nonlinearities from the cochlea, signal processing local to the midbrain, and any top-down processing due to cortical/perceptual modulations. We can rule out explanations due to loudness differences (Supplemental Fig. S4), as all tokens were matched in sound level and perceptual loudness (94.1 ± 1.0 phon) (Moore et al., 1997). The patterns observed in the FFR are also unlikely attributed to acoustical F0 or F1 spectral amplitude differences. In fact, an acoustic analysis showed a linear declining pattern in the acoustic F0 amplitude and invariance in F1 (Fig.

S4). Cochlear processing leading to the FFR can be highly nonlinear and as such, FFRs might not be expected to exactly mirror the acoustic stimulus input (Bidelman and Bhagat, 2020). However, simulated FFRs from an auditory nerve model, which reflect the output of important cochlear transformations (but not attention or perceptual processing), failed to predict the empirical data and instead more closely resembled acoustic changes in the stimuli (Fig. S5). While qualitative, these results are in stark contrast to the patterns observed in neural FFRs, which showed enhancements for endpoint tokens at both sides of the continuum vs. the midpoint. In this regard, our findings that speech-FFRs to otherwise identical speech tokens of the continuum were (i) enhanced during active categorization compared to their passively-evoked counterparts, (ii) show a unique response profile compared to their acoustic or cochlear counterparts, and (iii) were warped in the direction of listeners' phonetic labeling provides strong evidence for a neuro-perceptual origin of our data.

Our findings are reminiscent but contrast results of Chandrasekaran et al. (2009), who showed FFRs to an otherwise identical speech stimulus (/da/) presented in a variable vs. predictable order led to differences in F1 (but not F0 encoding). Several differences in study design may account for these divergent findings. Cursory differences in their CV vs. our vowel stimuli aside, context in Chandrasekaran et al. (2009) was manipulated at the between-token level whereas we varied context on a higher-order scale (between trains). Moreover, serial presentation in our forward and backward conditions leads to strong predictive hearing as listeners can fully anticipate subsequent tokens along the continuum. Contrastively, the predictable nature of stimuli in Chandrasekaran et al. (2009) was achieved via the repetition of a single token. Third, their use of alternating vs. single polarity stimulus presentation (as used here) can lead to a differential weighting of F0 vs. F1 in the speech-FFR (Kumar et al., 2014), which may alter context-dependent effects in a frequency-dependent manner. Perhaps more critically, the context-dependence observed in Chandrasekaran et al. (2009) was observed under passive listening, so it is not clear how those findings relate to the perception of the eliciting speech stimuli vs. other, more generalized, auditory mechanisms, *per se* (e.g., stimulus-specific adaptation) (Pérez-González et al., 2005). In contrast, speech-FFRs in the present study were recorded during an active speech identification task which led to context-order effects at both F0 and F1.

Two alternative theories suggest how ambiguous phonemes are categorized and might account for category-level coding we find in FFRs: the Natural Referent Vowel (NRV) and the Native Language Magnet (NLM) models. The NRV proposes that spectral prominences that are easier to detect lead to directional asymmetries in category discrimination tasks. Contrastively, the NLM proposes that directional asymmetries are caused by the vowel space being biased towards native phonetic prototypes—built through long-term statistical learning— which act as perceptual magnets for ambiguous phonemes (Masapollo et al., 2017; Zhao et al., 2019). Our findings have support from both models, but more strongly support the NLM within the context of categorization. Perception of ambiguous tokens was driven by listeners categorizing the sound as one of the prototypical tokens (i.e., NLM). This interpretation is further bolstered by the U shape found in FFR responses, which suggests there is pull from both prototypes. However, the fact Tk4 responses were more strongly correlated with $Tk1_{FOR}$ than $Tk7_{BACK}$ responses suggests there may be a slight bias towards vowels with more prominent F0/F1 configuration, consistent with NRV. It could be that the vowel biasing is more strongly related to NLM in our study as we used a continuum with prototypes that were both native to our listeners. Previous studies supporting NRV interpretations of FFR and direction-dependent vowel coding effects compared within-category stimuli (Masapollo et al., 2017; Zhao et al., 2019).

Additional evidence that the FFR reflects aspects of speech percepts was our finding that response components were associated with listeners' categorical boundary and the slope of their psychometric function. We and others have shown that perception begins to differentiate phonemes categorially early in the cortical hierarchy and no later than primary auditory cortex (Bidelman and Lee, 2015; Bidelman and Walker, 2019; Carter and Bidelman, 2021; Chang et al., 2010). Here, we extend these findings by showing category-specific neural representations extend as low as the brainstem. As the FFR is largely driven by midbrain regions (Bidelman, 2018b), the link between FFR and psychometric measures is consistent with notions that low-level auditory representations carry information regarding signal clarity, strength of categorization, and vowel identity (Binder et al., 2004). In contrast, FFR measures did not predict the speed of listeners' decisions. RTs are however largely driven by higher-order frontal brain regions (Binder et al., 2004), so it is perhaps not surprising that FFRs failed to predict perceptual speeds (but see Galbraith et al. 2000). Interestingly, while categorization in FFRs differentiated along the response *amplitude,* the response *frequency* accounted for variability in the brain-behavior relationships (see Tables S1, S2). The explanation for this dichotomy in driving different aspects of categorical behaviors is unclear. But it is conceivable that FFR amplitude might relate more strongly with the underlying strength of the categorical percept whereas frequency may relate more to the movement of those percepts along the acoustic dimension producing warping.

Collectively, the fact that FFRs are strongly modulated by attention and show category-specificity strongly suggests brainstem FFRs carry perceptual correlates related to how listeners ultimately hear the speech signal. Broadly, our findings indicate top-down processes modulate brainstem representations to fit the anticipated speech percept. Our data bolster notions that the FFR carries perceptually relevant cues related to phonetic representations and is thus more than just a neural mirror of the acoustic signal. Together, our findings suggest that midbrain plays a vital role in the active perception and categorization of speech.

Code data availability statement

Processed data are included as Supplemental Materials. Raw data and code supporting the conclusions of this article will be made available by the authors, without undue reservation.

Data availability

See code and data availability statement.

Credit authorship contribution statement

Jared A. Carter: Investigation, Data curation, Formal analysis, Writing – original draft. **Gavin M. Bidelman:** Investigation, Formal analysis, Writing – original draft.

Acknowledgments

Work supported by the National Institute on Deafness and Other Communication Disorders (R01DC016267). Requests for data and materials should be directed to G.M.B. [gbidel@indiana.edu].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.119899.

References

Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. Hear. Res. 245, 35–47.

Alho, J., Green, B.M., May, P.J., Sams, M., Tiitinen, H., Rauschecker, J.P., Jääskeläinen, I.P., 2016. Early-latency categorical speech sound representations in the left inferior frontal gyrus. Neuroimage 129, 214–223.

Altmann, C.F., Uesaki, M., Ono, K., Matsuhashi, M., Mima, T., Fukuyama, H., 2014. Categorical speech perception during active discrimination of consonants and vowels. Neuropsychologia 64, 13–23.
Anderson, S., Parbery-Clark, A., White-Schwoch, T., Kraus, N., 2012. Aging affects neural precision of speech encoding. J. Neurosci. 32, 14156–14164.
Asilador, A., Llano, D.A., 2021. Top-down inference in the auditory system: potential roles for corticofugal projections. Front. Neural Circuits 14.
Bajo, V.M., Nodal, F.R., Moore, D.R., King, A.J., 2010. The descending corticocollicular pathway mediates learning-induced auditory plasticity. Nat. Neurosci. 13, 253–260.
Bathellier, B., Ushakova, L., Rumpel, S., 2012. Discrete neocortical dynamics predict behavioral categorization of sounds. Neuron 76, 435–449.
Beddor, P.S., Harnsberger, J.D., Lindemann, S., 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. J. Phon. 30, 591–627.
Bidelman, G., 2013. The role of the auditory brainstem in processing musically relevant pitch. Front. Psychol. 4, 1–13.
Bidelman, G., Powers, L., 2018. Response properties of the human frequency-following response (FFR) to speech and non-speech sounds: level dependence, adaptation and phase-locking limits. Int. J. Audiol. 57, 665–672.
Bidelman, G.M., 2014. Objective information-theoretic algorithm for detecting brainstem evoked responses to complex stimuli. J. Am. Acad. Audiol. 25, 711–722.
Bidelman, G.M., 2015a. Induced neural beta oscillations predict categorical speech perception abilities. Brain Lang. 141, 62–69.
Bidelman, G.M., 2015b. Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. Hear. Res. 323, 68–80.
Bidelman, G.M., 2015c. Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. J. Neurosci. Methods 241, 94–100.
Bidelman, G.M., 2016. Relative contribution of envelope and fine structure to the subcortical encoding of noise-degraded speech. J. Acoust. Soc. Am. 140, EL358–EL363.
Bidelman, G.M., 2018a. Sonification of scalp-recorded frequency-following responses (FFRs) offers improved response detection over conventional statistical metrics. J. Neurosci. Methods 293, 59–66.
Bidelman, G.M., 2018b. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. Neuroimage 175, 56–69.
Bidelman, G.M., Bhagat, S., 2020. Brainstem correlates of cochlear nonlinearity measured via the scalp-recorded frequency-following response. Neuroreport 31, 702–707.
Bidelman, G.M., Lee, C.C., 2015. Effects of language experience and stimulus context on the neural organization and categorical perception of speech. Neuroimage 120, 191–200.
Bidelman, G.M., Momtaz, S., 2021. Subcortical rather than cortical sources of the frequency-following response (FFR) relate to speech-in-noise perception in normal-hearing listeners. Neurosci. Lett. 746, 135664.
Bidelman, G.M., Moreno, S., Alain, C., 2013. Tracing the emergence of categorical perception in the human auditory system. Neuroimage 29, 201–212.
Bidelman, G.M., Pearson, C., Harrison, A., 2021. Lexical influences on categorical speech perception are driven by a temporoparietal circuit. J. Cogn. Neurosci. 33, 840–852.
Bidelman, G.M., Price, C.N., Shen, D., Arnott, S.R., Alain, C., 2019. Afferent-efferent connectivity between auditory brainstem and cortex accounts for poorer speech-in-noise comprehension in older adults. Hear. Res. 382, 107795.
Bidelman, G.M., Villafuerte, J.W., Moreno, S., Alain, C., 2014a. Age-related changes in the subcortical-cortical encoding and categorical perception of speech. Neurobiol. Aging 35, 2526–2540.
Bidelman, G.M., Villafuerte, J.W., Moreno, S., Alain, C., 2014b. Age-related changes in the subcortical–cortical encoding and categorical perception of speech. Neurobiol. Aging 35, 2526–2540.
Bidelman, G.M., Walker, B., 2019. Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network. Neuroimage 201, 116022.
Bidelman, G.M., Walker, B.S., 2017. Attentional modulation and domain-specificity underlying the neural organization of auditory categorical perception. Eur. J. Neurosci. 45, 690–699.
Bidelman, G.M., Weiss, M.W., Moreno, S., Alain, C., 2014c. Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. Eur. J. Neurosci. 40, 2662–2673.
Billings, C.J., Bologna, W.J., Muralimanohar, R.K., Madsen, B.M., Molis, M.R., 2019. Frequency following responses to tone glides: Effects of frequency extent, direction, and electrode montage. Hear. Res. 375, 25–33.
Billings, C.J., Tremblay, K.L., Stecker, G.C., Tolin, W.M., 2009. Human evoked cortical activity to signal-to-noise ratio and absolute signal level. Hear. Res. 254, 15–24.
Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. 7, 295–301.
Bones, O., Hopkins, K., Krishnan, A., Plack, C.J., 2014. Phase locked neural activity in the human brainstem predicts preference for musical consonance. Neuropsychologia 58, 23–32.
Bradley, J.V., 1958. Complete counterbalancing of immediate sequential effects in a Latin square design. J. Am. Stat. Assoc. 53, 525–528.
Brugge, J.F., Nourski, K.V., Oya, H., Reale, R.A., Kawasaki, H., Steinschneider, M., Howard III, M.A., 2009. Coding of repetitive transients by auditory cortex on Heschl's gyrus. J. Neurophysiol. 102, 2358–2374.
Burghard, A., Voigt, M.B., Kral, A., Hubka, P., 2019. Categorical processing of fast temporal sequences in the guinea pig auditory brainstem. Commun. Biol. 2, 1–9.
Campbell, T., Kerlin, J.R., Bishop, C.W., Miller, L.M., 2012. Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. Ear Hear. 33, 144.
Carter, J., 2018. Informational and energetic masking effects on speech-evoked cortical auditory potentials. Department of Speech, Language, and Hearing Sciences. University of Arizona, Tucson, AZ.
Carter, J.A., Bidelman, G.M., 2021. Auditory cortex is susceptible to lexical influence as revealed by informational vs. energetic masking of speech categorization. Brain Res. 1759, 147385.
Carter, J.A., Buder, E.H., Bidelman, G.M., 2022. Nonlinear dynamics in auditory cortical activity reveal the neural basis of perceptual warping in speech categorization. JASA Express Lett. 2, 045201.
Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., Kraus, N., 2009. Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. Neuron 64, 311–319.
Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: Neural origins and plasticity. Psychophysiology 47, 236–246.
Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T., 2010. Categorical speech representation in human superior temporal gyrus. Nat. Neurosci. 13, 1428–1432.
Cheng, F.Y., Xu, C., Gold, L., Smith, S., 2021. Rapid enhancement of subcortical neural responses to sine-wave speech. Front. Neurosci. 15.
Coffey, E.B., Herholz, S.C., Chepesiuk, A.M., Baillet, S., Zatorre, R.J., 2016. Cortical contributions to the auditory frequency-following response revealed by MEG. Nat. Commun. 7, 1–11.
Coffey, E.B.J., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zatorre, R.J., Kraus, N., 2019. Evolving perspectives on the sources of the frequency–following response. Nat. Commun. 10, 5036.
Diehl, R.L., Elman, J.L., McCusker, S.B., 1978. Contrast effects on stop consonant identification. J. Exp. Psychol. Hum. Percept. Perform. 4, 599.
Dunlop, C., Webster, W., Simons, L., 1965. Effect of attention on evoked responses in the classical auditory pathway. Nature 206, 1048–1050.
Eimas, P.D., Corbit, J.D., 1973. Selective adaptation of linguistic feature detectors. Cognit. Psychol. 4, 99–109.
Forte, A.E., Etard, O., Reichenbach, T., 2017. The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. eLife 6, e27203.
Galbraith, G.C., Arbagey, P.W., Branski, R., Comerci, N., Rector, P.M., 1995. Intelligible speech encoded in the human brain stem frequency-following response. Neuroreport 6, 2363–2367.
Galbraith, G.C., Bhuta, S.M., Choate, A.K., Kitahara, J.M., Mullen Jr, T.A., 1998. Brain stem frequency-following response to dichotic vowels during attention. Neuroreport 9, 1889–1893.
Galbraith, G.C., Brown, W.S., 1990. Cross-correlation and latency compensation analysis of click-evoked and frequency-following brain-stem responses in man. Electroencephalogr. Clin. Neurophysiol. Evoked Potentials Sect. 77, 295–308.
Galbraith, G.C., Chae, B.C., Cooper, J.R., Gindi, M.M., Ho, T.N., Kim, B.S., Mankowski, D.A., Lunde, S.E., 2000. Brainstem frequency-following response and simple motor reaction time. Int. J. Psychophysiol. 36, 35–44.
Galbraith, G.C., Kane, J.M., 1993. Brainstem frequency-following responses and cortical event-related potentials during attention. Percept. Mot. Skills 76, 1231–1241.
Ganong III, W.F., Zatorre, R.J., 1980. Measuring phoneme boundaries four ways. J. Acoust. Soc. Am. 68, 431–439.
Gardi, J., Merzenich, M., McKean, C., 1979. Origins of the scalp-recorded frequency-following response in the cat. Audiology 18, 353–380.
Gnanateja, G.N., Rupp, K., Llanos, F., Remick, M., Pernia, M., Sadagopan, S., Teichert, T., Abel, T.J., Chandrasekaran, B., 2021. Frequency-following responses to speech sounds are highly conserved across species and contain cortical contributions. Eneuro 8.
Gockel, H.E., Muhammed, L., Farooq, R., Plack, C.J., Carlyon, R.P., 2013. No evidence for ITD-specific adaptation in the frequency following response. In: Basic Aspects of Hearing. Springer, pp. 231–238.
Goldstone, R.L., Hendrickson, A.T., 2010. Categorical perception. Wiley Interdiscip. Rev. Cogn. Sci. 1, 69–78.
Gorina-Careta, N., Kurkela, J.L.O., Hämäläinen, J., Astikainen, P., Escera, C., 2021. Neural generators of the frequency-following response elicited to stimuli of low and high frequency: A magnetoencephalographic (MEG) study. Neuroimage 231, 117866.
Guo, N., Si, X., Zhang, Y., Ding, Y., Zhou, W., Zhang, D.R., Hong, B., 2021. Speech frequency-following response in human auditory cortex is more than a simple tracking. Neuroimage 226, 117545.
Hansen, T., Olkkonen, M., Walter, S., Gegenfurtner, K.R., 2006. Memory modulates color appearance. Nat. Neurosci. 9, 1367–1368.
Harris, K.C., Wilson, S., Eckert, M.A., Dubno, J.R., 2012. Human evoked cortical activity to silent gaps in noise: effects of age, attention, and cortical processing speed. Ear Hear. 33, 330–339.
Hartmann, T., Weisz, N., 2019. Auditory cortical generators of the Frequency Following Response are modulated by intermodal attention. Neuroimage 203, 116185.
Healy, A.F., Repp, B.H., 1982. Context independence and phonetic mediation in categorical perception. J. Exp. Psychol. Hum. Percept. Perform. 8, 68.
Hillyard, S.A., Hink, R.F., Schwent, V.L., Picton, T.W., 1973. Electrical signs of selective attention in the human brain. Science 182, 177–180.
Hillyard, S.A., Picton, T.W., 1979. Event-related brain potentials and selective information processing in man. In: Desmedt, J.E. (Ed.), Progress in Clinical Neurophysiology. Karger, Basel, pp. 1–52.
Holmes, E., Purcell, D.W., Carlyon, R.P., Gockel, H.E., Johnsrude, I.S., 2018. Attentional modulation of envelope-following responses at lower (93–109 Hz) but not higher (217–233 Hz) modulation rates. J. Assoc. Res. Otolaryngol. 19, 83–97.
Johnson, K.L., Nicol, T.G., Kraus, N., 2005. Brain stem response to speech: a biological marker of auditory processing. Ear Hear. 26, 424–434.

Joris, P., Schreiner, C., Rees, A., 2004. Neural processing of amplitude-modulated sounds. Physiol. Rev. 84, 541–577.
Krishnan, A., 2002. Human frequency-following responses: representation of steady-state synthetic vowels. Hear. Res. 166, 192–201.
Krishnan, A., Gandour, J.T., Ananthakrishnan, S., Bidelman, G.M., Smalt, C.J., 2011. Linguistic status of timbre influences pitch encoding in the brainstem. Neuroreport 22, 801–803.
Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010. The effects of tone language experience on pitch processing in the brainstem. J. Neurolinguist. 23, 81–95.
Krishnan, A., Gandour, J.T., Bidelman, G.M., 2012. Experience-dependent plasticity in pitch encoding: from brainstem to auditory cortex. Neuroreport 23, 498.
Krishnan, A., Gandour, J.T., Bidelman, G.M., Swaminathan, J., 2009. Experience dependent neural representation of dynamic pitch in the brainstem. Neuroreport 20, 408.
Krizman, J., Bonacina, S., Otto-Meyer, R., Kraus, N., 2021a. Non-stimulus-evoked activity as a measure of neural noise in the frequency-following response. J. Neurosci. Methods 362, 109290.
Krizman, J., Tierney, A., Nicol, T., Kraus, N., 2021b. Listening in the moment: how bilingualism interacts with task demands to shape active listening. Front. Neurosci. 15.
Kuhl, P.K., 1986. Theoretical contributions of tests on animals to the special-mechanisms debate in speech. Exp. Biol. 45, 233–265.
Kuhl, P.K., Miller, J.D., 1975. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. Science 190, 69–72.
Kumar, K., Bhat, J.S., D'Costa, P.E., Srivastava, M., Kalaiah, M.K., 2014. Effect of stimulus polarity on speech evoked auditory brainstem response. Audiol. Res. 3, e8 -e8.
Lai, J., Price, C.N., Bidelman, G.M., 2022. Brainstem speech encoding is dynamically shaped online by fluctuations in cortical α state. Neuroimage 263, 119627.
Langner, G., Schreiner, C.E., 1988. Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J. Neurophysiol. 60, 1799–1822.
Lesicko, A.M., Geffen, M.N., 2022. Diverse functions of the auditory cortico-collicular pathway. Hear. Res., 108488.
Liu, D., Hu, J., Dong, R., Chen, J., Musacchia, G., Wang, S., 2018. Effects of inter-stimulus interval on speech-evoked frequency-following response in elderly adults. Front. Aging Neurosci. 10.
Liu, M., Xie, F., Dai, J., Zhang, J., Yuan, K., Wang, N., 2023. Brain-wide inputs to the non-lemniscal inferior colliculus in mice. Neurosci. Lett. 793, 136976.
López-Caballero, F., Martin-Trias, P., Ribas-Prats, T., Gorina-Careta, N., Bartrés-Faz, D., Escera, C., 2020. Effects of cTBS on the frequency-following response and other auditory evoked potentials. Front. Hum. Neurosc. 14.
Luo, H., Boemio, A., Gordon, M., Poeppel, D., 2007. The perception of FM sweeps by Chinese and English listeners. Hear. Res. 224, 75–83.
Mahmud, M.S., Yeasin, M., Bidelman, G.M., 2021. Data-driven machine learning models for decoding speech categorization from evoked brain responses. J. Neural Eng. 18, 046012.
Mankel, K., Barber, J., Bidelman, G.M., 2020. Auditory categorical processing for speech is modulated by inherent musical listening skills. Neuroreport 31, 162.
Mankel, K., Bidelman, G.M., 2018. Inherent auditory skills rather than formal music training shape the neural encoding of speech. Proc. Natl Acad. Sci. 115, 13129–13134.
Masapollo, M., Polka, L., Molnar, M., Ménard, L., 2017. Directional asymmetries reveal a universal bias in adult vowel perception. J. Acoust. Soc. Am. 141, 2857–2869.
Moore, B.C., Glasberg, B.R., Baer, T., 1997. A model for the prediction of thresholds, loudness, and partial loudness. J. Audio Eng. Soc. 45, 224–240.
Musacchia, G., Strait, D., Kraus, N., 2008. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear. Res. 241, 34–42.
Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. 118, 2544–2590.
Oatman, L., Anderson, B., 1980. Suppression of the auditory frequency following response during visual attention. Electroencephalogr. Clin. Neurophysiol. 49, 314–322.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Olthof, B.M.J., Rees, A., Gartside, S.E., 2019. Multiple nonauditory cortical regions innervate the auditory midbrain. J. Neurosci. 39, 8916–8928.
Pérez-González, D., Malmierca, M., 2014. Adaptation in the auditory system: an overview. Front. Integr. Neurosci. 8.
Pérez-González, D., Malmierca, M.S., Covey, E., 2005. Novelty detector neurons in the mammalian auditory midbrain. Eur. J. Neurosci. 22, 2879–2885.
Pisoni, D.B., 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Percept. Psychophys. 13, 253–260.
Pisoni, D.B., Tash, J., 1974. Reaction times to comparisons within and across phonetic categories. Percept. Psychophys. 15, 285–290.
Price, C.N., Bidelman, G.M., 2021. Attention reinforces human corticofugal system to aid speech perception in noise. Neuroimage 235, 118014.
Reetzke, R., Xie, Z., Llanos, F., Chandrasekaran, B., 2018. Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood. Curr. Biol. 28, 1419–1427.
Ross, B., Tremblay, K.L., Alain, C., 2020. Simultaneous EEG and MEG recordings reveal vocal pitch elicited cortical gamma oscillations in young and older adults. Neuroimage 204, 116253.
Russo, N., Nicol, T., Musacchia, G., Kraus, N., 2004. Brainstem responses to speech syllables. Clin. Neurophysiol. 115, 2021–2030.
Schouten, M.E.H., 1985. Identification and discrimination of sweep tones. Percept. Psychophys. 37, 369–376.
Shukla, B., Bidelman, G.M., 2021. Enhanced brainstem phase-locking in low-level noise reveals stochastic resonance in the frequency-following response (FFR). Brain Res. 1771, 147643.
Skoe, E., Kraus, N., 2010. Auditory brainstem response to complex sounds: a tutorial. Ear Hear. 31, 302.
Slugocki, C., Bosnyak, D., Trainor, L.J., 2017. Simultaneously-evoked auditory potentials (SEAP): A new method for concurrent measurement of cortical and subcortical auditory-evoked activity. Hear. Res. 345, 30–42.
Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. Electroencephalogr. Clin. Neurophysiol. 39, 465–472.
Sohmer, H., Pratt, H., Kinarti, R., 1977. Sources of frequency following responses (FFR) in man. Electroencephalogr. Clin. Neurophysiol. 42, 656–664.
Strouse, A., Ashmead, D.H., Ohde, R.N., Grantham, D.W., 1998. Temporal processing in the aging auditory system. J. Acoust. Soc. Am. 104, 2385–2399.
Suga, N., 2008. Role of corticofugal feedback in hearing. J. Comp. Physiol. A 194, 169–183.
Suga, N., Gao, E., Zhang, Y., Ma, X., Olsen, J.F., 2000. The corticofugal system for hearing: recent progress. Proc. Natl Acad. Sci. 97, 11807–11814.
Sumner, M., 2011. The role of variation in the perception of accented speech. Cognition 119, 131–136.
Tang, H., Crain, S., Johnson, B.W., 2016. Dual temporal encoding mechanisms in human auditory cortex: evidence from MEG and EEG. Neuroimage 128, 32–43.
Tichko, P., Skoe, E., 2017. Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators. Hear. Res. 348, 1–15.
Tuller, B., Case, P., Ding, M., Kelso, J.A.S., 1994. The nonlinear dynamics of speech categorization. J. Exp. Psychol. Hum. Percept. Perform. 20, 3–16.
Tuller, B., Jantzen, M.G., Jirsa, V.K., 2008. A dynamical approach to speech categorization: two routes to learning. New Ideas Psychol. 26, 208–226.
Varghese, L., Bharadwaj, H.M., Shinn-Cunningham, B.G., 2015. Evidence against attentional state modulating scalp-recorded auditory brainstem steady-state responses. Brain Res. 1626, 146–164.
Wallace, M.N., Rutkowski, R.G., Shackleton, T.M., Palmer, A.R., 2000. Phase-locked responses to pure tones in guinea pig auditory cortex. Neuroreport 11, 3989–3993.
Winkler, I., Reinikainen, K., Näätänen, R., 1993. Event-related brain potentials reflect traces of echoic memory in humans. Percept. Psychophys. 53, 443–449.
Xie, Z., Reetzke, R., Chandrasekaran, B., 2019. Machine learning approaches to analyze speech-evoked neurophysiological responses. J. Speech Lang. Hear. Res. 62, 587–601.
Xu, Y., Gandour, J.T., Francis, A.L., 2006a. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. J. Acoust. Soc. Am. 120, 1063–1074.
Xu, Y., Krishnan, A., Gandour, J.T., 2006b. Specificity of experience-dependent pitch representation in the brainstem. Neuroreport 17, 1601–1605.
Yellamsetty, A., Bidelman, G.M., 2019. Brainstem correlates of concurrent speech identification in adverse listening conditions. Brain Res. 1714, 182–192.
Yi, H.G., Xie, Z., Reetzke, R., Dimakis, A.G., Chandrasekaran, B., 2017. Vowel decoding from single-trial speech-evoked electrophysiological responses: A feature-based machine learning approach. Brain Behav. 7, e00665.
Zhang, C., Lu, L., Wu, X., Li, L., 2014. Attentional modulation of the early cortical representation of speech signals in informational or energetic masking. Brain Lang. 135, 85–95.
Zhao, T.C., Masapollo, M., Polka, L., Ménard, L., Kuhl, P.K., 2019. Effects of formant proximity and stimulus prototypicality on the neural discrimination of vowels: Evidence from the auditory frequency-following response. Brain Lang. 194, 77–83.
Zilany, M.S.A., Bruce, I.C., Carney, L.H., 2014. Updated parameters and expanded simulation options for a model of the auditory periphery. J. Acoust. Soc. Am. 135, 283–286.

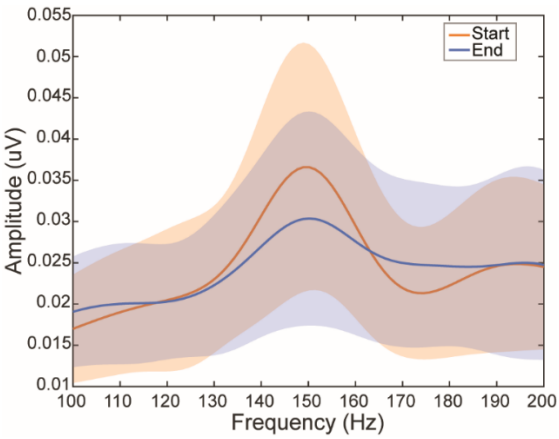

Figure S1. Brainstem FFRs show insignificant adaptation in response to rapid speech trains. Shown here are FFR spectra around the F0 frequency for the start and end tokens of the stimulus trains (pooling tokens and conditions). Shaded areas = ±1 s.e.m.

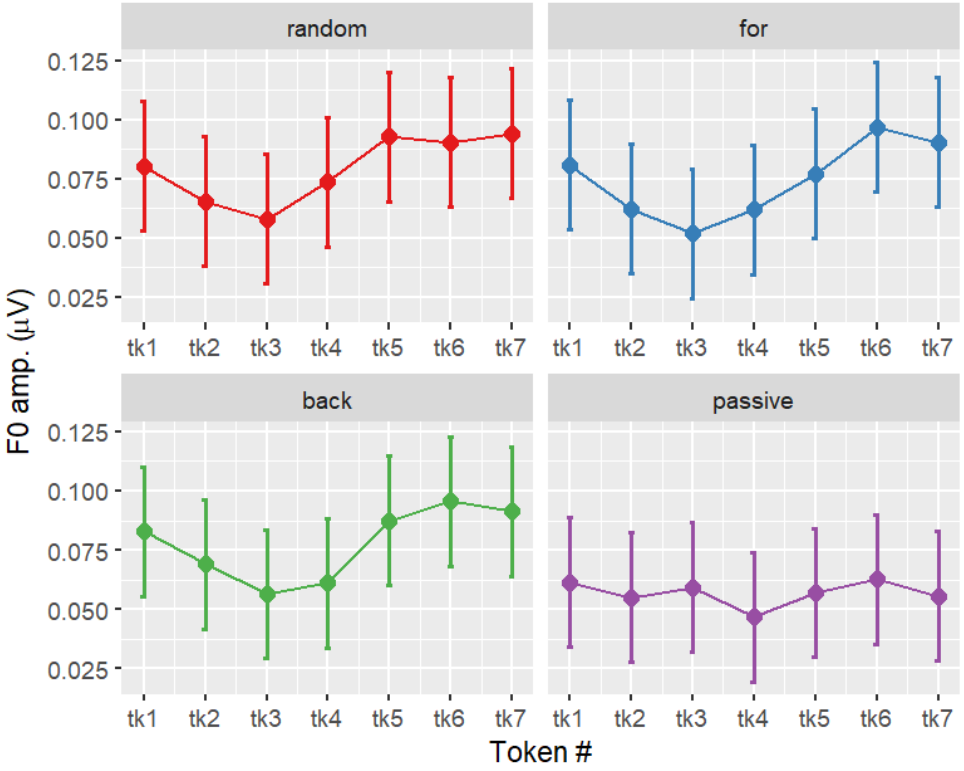

Figure S2. F0 amplitudes showed a significant quadratic trend (U-shape) in forward and backward conditions. Despite an appearance, the quadratic pattern was not significant for the random and passive conditions. Error bars = ±95% CI.

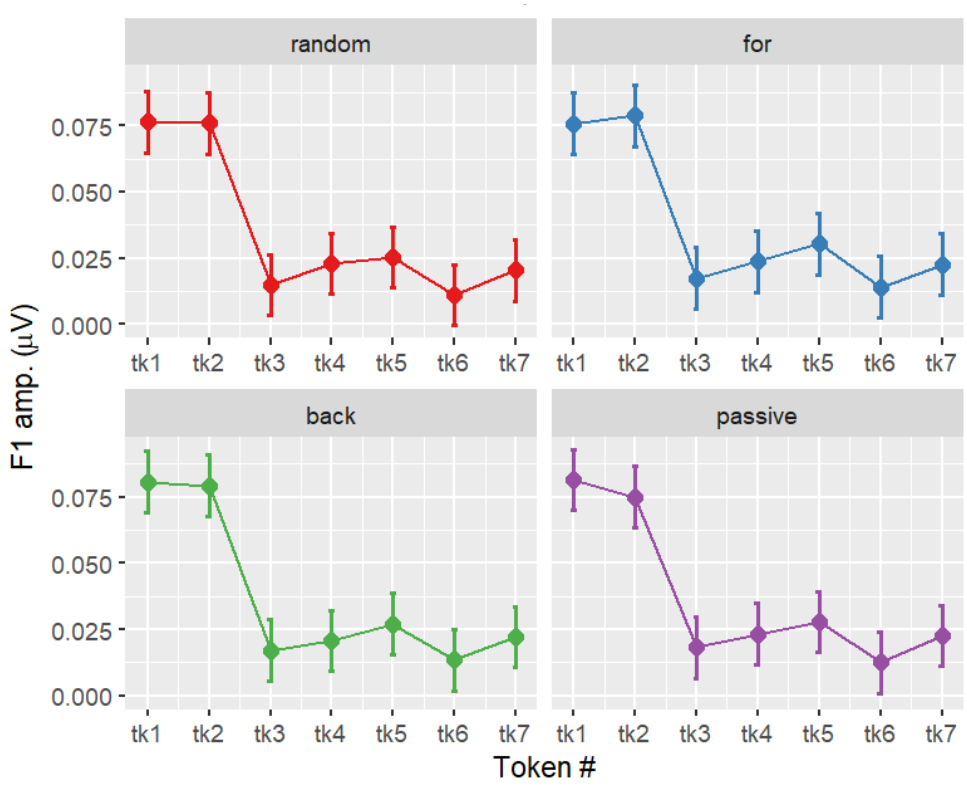

Figure S3. F1 amplitudes showed a significant quadratic trend (U-shape) in all four conditions. Error bars = ±95% CI.

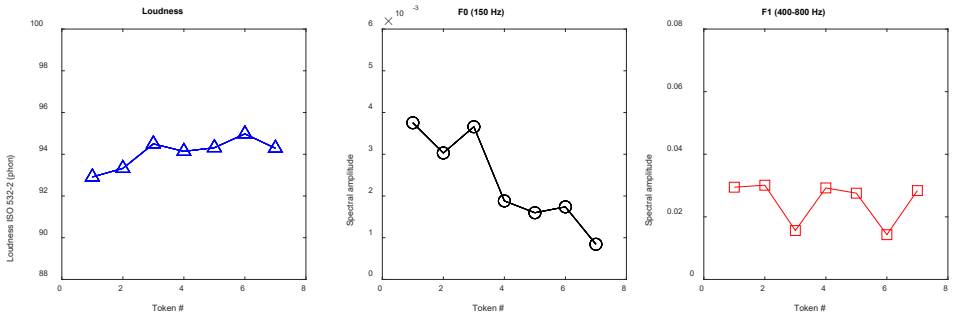

Figure S4. *Acoustic* properties of the stimulus tokens. (***left***) Loudness computed using the model of Moore et al. (1997) based on the ISO 532.2 standard. Loudness varies minimally (< 1 phon) across tokens. (***middle***) F0 spectral amplitudes. Note the declining linear trend in F0 as compared to the quadratic (categorical) U-shape observed in FFR data (see Fig. S2). (***right***) F1 amplitudes. Acoustically, F1 is stronger than F0 owing to its boost near the formant resonance. However, unlike in FFRs, F1 varies little across the continuum.

Simulated FFRs from a computational AN model

We used a computational model of the auditory nerve (AN) (Zilany et al., 2009; Zilany and Carney, 2010; Zilany et al., 2014) to simulate brainstem FFRs (Bidelman, 2014; Dau, 2003). Details of this phenomenological model and implementation to model FFRs are provided in Bidelman (2014). The model incorporates several important nonlinearities observed in the auditory periphery, including cochlear filtering, level-dependent gain (i.e., compression) and bandwidth control, long-term adaptation, as well as two-tone suppression. Model tuning curves were fit to the characteristic frequency (CF)-dependent variation in threshold and bandwidth for high-spontaneous rate (SR) fibers in normal-hearing cats (Miller et al., 1997). The stochastic nature of AN responses is accounted for by a modified non-homogenous Poisson process, which includes effects of both absolute and relative refractory periods and captures the major stochastic properties of single-unit AN responses (e.g., Young and Barta, 1986).

The AN model was used to simulate the scalp-recorded FFR using methodology described by Dau (2003) and detailed in Bidelman (2014) (**Fig. S5A**). This approach is based on the assumption that the far-field FFR recorded at the scalp is a convolution of an elementary unit waveform (i.e., impulse response) with the instantaneous discharge rate from a given auditory nucleus (Dau, 2003; Goldstein and Kiang, 1958).

We submitted 50 repetitions of each vowel to the model to evoke AN spike-trains. Spikes were generated from each of 100 model fibers (CFs: 125-11000 Hz; high spontaneous rate units) to simulate the discharge pattern across the cochlear partition. Activity from the entire ensemble was then summed to form a population post-stimulus time histogram (PSTH). The PSTH was then convolved with a unitary response function, simulating the impulse response of nuclei from the auditory brainstem (for details, see Dau, 2003). Finally, pink noise (1/*f* distribution) was added to simulate the quasi-stochastic nature of EEG noise (Bidelman, 2014; Dau, 2003; Granzow et al., 2001). Resulting model waveforms provided a mirror approximation of the time-frequency characteristics of true FFRs recorded in our human listeners (**Fig. S5B**,**C**). As with the empirical FFR recordings, we measured model F0 (150 Hz) and F1 (400-800 Hz) amplitudes from response spectra. This allowed us to qualitatively compare true FFRs (recorded during an active perceptual warping task) with model responses, which similarly reflect the output of cochlear processing (e.g., spectral decomposition, nonlinearities) but are not subject to attention, perception, and/or higher-level cortical processing as in the empirical recordings.

Figure S5 shows model FFRs to tokens along the vowel continuum. In general, model response F0 and F1 amplitudes followed a similar pattern to the F0 and F1 trends of the acoustic stimuli (cf. Fig. S4). Critically, model FFRs did not show enhancements near category endpoints (i.e., U-shape) as in the empirical FFRs (e.g., Figs. 7-8). These findings suggest, at least qualitatively, that category coding effects observed in the FFR are not due to stimulus acoustics or cochlear nonlinearities, *per se*, but instead reflect top-down modulations from listeners' perceptual state and attention.

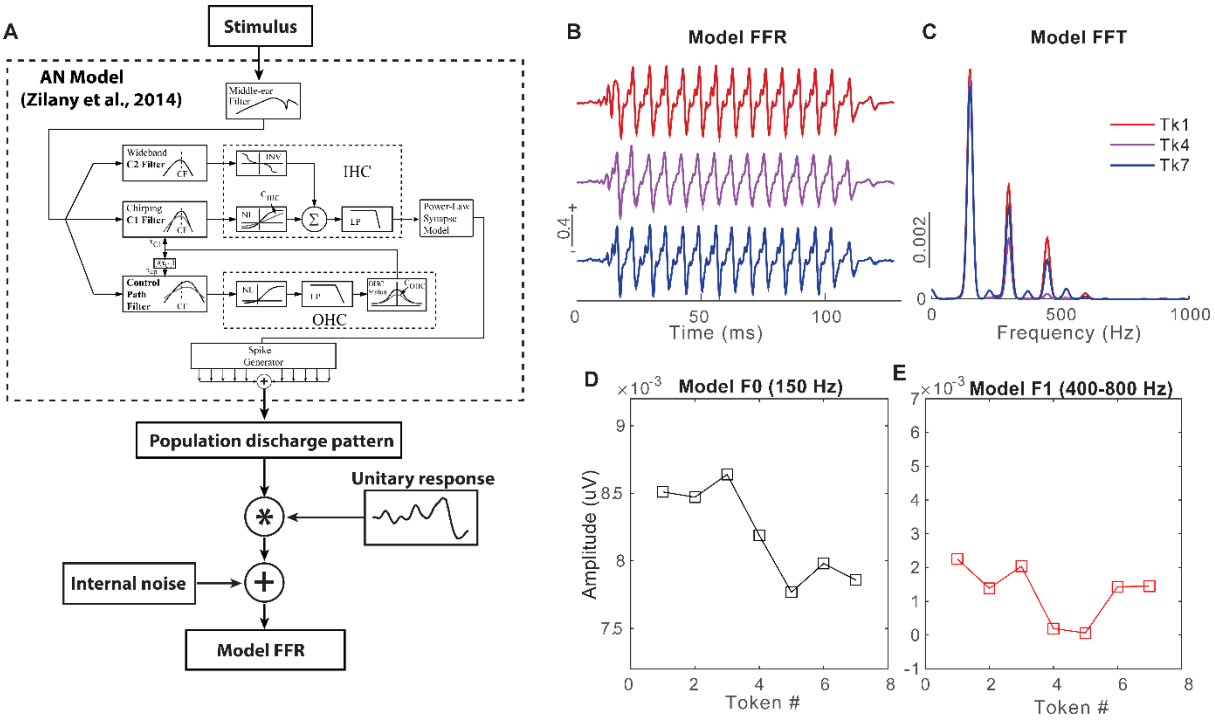

Figure S5. Computational model architecture used to simulate scalp-recorded FFRs (Bidelman, 2014). (**A**) The acoustic stimulus is input to a biologically plausible model of the auditory periphery (Zilany et al., 2014). The model provides a simulated realization of the neural discharge pattern for single AN fibers. After middle-ear filtering and hair cell transduction and filtering, action potentials are generated according to a nonhomogeneous Poisson process. Spikes were generated from 100 model fibers (CFs: 125-11000 Hz) to simulate neural activity across the cochlear partition and summed to form a population PSTH for the entire AN array. Population PSTHs were then convolved with a unitary response function which simulates the impulse response of nuclei within the auditory brainstem (Dau, 2003). Additive noise simulated the inherent random fluctuations in scalp-recorded EEG. (**B,C**) Model FFR time waveforms and response spectra. (**D,E**) Model F0 and F1 amplitudes follow a similar trend as the acoustic F0/F1 (cf. Fig. S4) but do not show the categorical pattern as observed in true FFRs (cf. Fig. 7, main text).

Table S1. ***GLME model fit parameters for predicting categorical boundary (β0) location from FFRs measures.***

Name	*Estimate*	*SE*	*t-stat*	*DF*	*p-value*	*Lower*	*Upper*
Intercept	0.16	0.18	0.93	11	0.37	-0.22	0.55
F0 Amp	-15.33	16.29	0.94	11	0.36	-51.19	20.52
F0 Freq	-0.003	0.005	0.59	11	0.57	-0.01	0.009
F1 Amp	-474.42	454.37	1.04	11	0.32	-1474.5	525.64
F1 Freq*	0.010	0.004	2.43	11	**0.033***	0.0009	0.018

Coefficients and significance tests for individual predictor variables from neural responses on the categorical boundary. *$p < 0.05$

Table S2. ***GLME model fit parameters for predicting psychometric slope (β1) from FFRs measures.***

Name	*Estimate*	*SE*	*t-stat*	*DF*	*p-value*	*Lower*	*Upper*
Intercept	-0.04	0.06	-0.65	11	0.53	-0.18	0.10
F0 Amp	-2.04	5.77	-0.35	11	0.73	-14.74	10.66
F0 Freq*	0.01	0.001	2.87	11	**0.015***	0.001	0.01
F1 Amp	329.63	160.96	2.05	11	0.065	-24.65	683.9
F1 Freq	-0.001	0.001	-0.99	11	0.34	-0.005	0.002

Coefficients and significance tests for individual predictor variables from neural responses on the psychometric slope. *$p < 0.05$

References

Bidelman, G.M., 2014. Objective information-theoretic algorithm for detecting brainstem evoked responses to complex stimuli. Journal of the American Academy of Audiology 25, 711-722.

Dau, T., 2003. The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. Journal of the Acoustical Society of America 113, 936-950.

Goldstein, M.H., Kiang, N.Y.S., 1958. Synchrony of neural activity in electric responses evoked by transient acoustic stimuli. Journal of the Acoustical Society of America 30, 107–114.

Granzow, M., Riedel, H., Kollmeier, B., 2001. Single-sweep-based methods to improve the quality of auditory brain stem responses Part I: Optimized linear filtering. Z. Audiol. 40, 32–44.

Miller, R.L., Schilling, J.R., Franck, K.R., Young, E.D., 1997. Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers. Journal of the Acoustical Society of America 101, 3602-3616.

Young, E.D., Barta, P.E., 1986. Rate responses of auditory nerve fibers to tones in noise near masked threshold. Journal of the Acoustical Society of America 79, 426-442.

Zilany, M.S., Bruce, I.C., Nelson, P.C., Carney, L.H., 2009. A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. Journal of the Acoustical Society of America 126, 2390-2412.

Zilany, M.S., Carney, L.H., 2010. Power-law dynamics in an auditory-nerve model can account for neural adaptation to sound-level statistics. Journal of Neuroscience 30, 10380-10390.

Zilany, M.S.A., Bruce, I.C., Carney, L.H., 2014. Updated parameters and expanded simulation options for a model of the auditory periphery. Journal of the Acoustical Society of America 135, 283-286.