JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Nonlinear dynamics in auditory cortical activity reveal the neural basis of perceptual warping in speech categorization[a)]

Jared A. Carter,[1,b)] Eugene H. Buder,[2,c)] and Gavin M. Bidelman[3,c)]
[1]Institute for Intelligent Systems, University of Memphis, Memphis, Tennessee 38152, USA
[2]School of Communication Sciences and Disorders, University of Memphis, Memphis, Tennessee 38152, USA
[3]Department of Speech, Language and Hearing Sciences, Indiana University, , Bloomington, Indiana 47408, USA
jcrter29@memphis.edu, ehbuder@memphis.edu, gbidel@indiana.edu

Abstract: Surrounding context influences speech listening, resulting in dynamic shifts to category percepts. To examine its neural basis, event-related potentials (ERPs) were recorded during vowel identification with continua presented in random, forward, and backward orders to induce perceptual warping. Behaviorally, sequential order shifted individual listeners' categorical boundary, versus random delivery, revealing perceptual warping (biasing) of the heard phonetic category dependent on recent stimulus history. ERPs revealed later (~300 ms) activity localized to superior temporal and middle/inferior frontal gyri that predicted listeners' hysteresis/enhanced contrast magnitudes. Findings demonstrate that interactions between fronto-temporal brain regions govern top-down, stimulus history effects on speech categorization.

[Editor: Douglas D O'Shaughnessy] https://doi.org/10.1121/10.0009896

Received: 30 November 2021 **Accepted:** 3 March 2022 **Published Online:** 4 April 2022

1. Introduction

In speech perception, listeners group similar sensory cues to form discrete phonetic labels—the process of categorical perception (CP). Spectral features vary continuously. However, reducing acoustic cues to discrete categories enables more efficient use of speech sounds for linguistic processing.[1,2] The extent to which phonetic speech categories from acoustic-sensory cues are influenced by perceptual biasing (top-down influences) has been debated. On one hand, categories might arise due to innate psychophysiological constraints.[3] Alternatively, there is ample evidence that top-down processing influences speech categorization as suggested by enhancements observed in highly proficient listeners[4–7] and biasing effects, when individuals hear a different category depending on the surrounding speech context.[8]

Changes in auditory-perceptual categories due to stimulus history are a form of nonlinear dynamics. Nonlinear dynamics in CP are especially prominent at the perceptual boundary, where different patterns of behavioral identification can result for otherwise identical speech sounds: hysteresis (i.e., percept continuing in the same category beyond the theoretical boundary) or enhanced contrast (i.e., percept changing to the other category before the theoretical boundary).[9–11] Both stop consonant and vowel continua can produce context-dependent shifts in perception, though stronger perceptual warping occurs with more ambiguous speech sounds.[12]

Event-related brain potentials (ERPs) have been used to examine the neural underpinnings of speech categorization.[13–15] ERPs reveal the brain performs its acoustic-to-phonetic conversion within ~150 ms and differentiates even the same speech sounds when categorized with different perceptual labels.[13] Yet it remains unknown how neural representations of categories change with recent state history as seen in hysteresis and other perceptual nonlinearities inherent to speech perception.[10] Shifting percepts near a categorical boundary due to presentation order (i.e., how stimuli are sequenced) should yield measurable neural signatures if speech perception is indeed warped dynamically.

Here, we evaluated the effects of nonlinear dynamics on speech categorization and its brain basis. We aimed to resolve whether perceptual hysteresis in CP occurs at early (i.e., auditory-sensory) or later (i.e., higher-order, linguistic) stages of speech analysis. We measured behavioral and multichannel EEG responses during rapid phoneme identification tasks where tokens along an identical continuum were presented in random versus serial (forward or backward) order. Based on previous

[a)]Invited paper.
[b)]Also at: School of Communication Sciences and Disorders, University of Memphis, Memphis, TN 38152, USA. Author to whom correspondence should be addressed.
[c)]Also at: Institute for Intelligent Systems, University of Memphis, Memphis, TN 38152, USA.

studies examining nonlinear dynamics[9,10] and top-down influences in speech CP,[4,5,7] we hypothesized (1) the location of listeners' perceptual boundary would shift according to the direction of stimulus presentation (i.e., random versus forward versus backward) and (2) perceptual warping would be accompanied by late modulations in the ERPs.

2. Materials and methods

2.1 Participants

The sample included $N = 15$ young participants (23.3 ± 3.9 years; 5 females) averaging 16.7 ± 3.4 years of education. All spoke American English, had normal hearing (≤20 dB HL; 250–8000 Hz), minimal musical training (≤3 years; average = 1.0 ± 1.3 years), and were mostly right-handed (mean = 75% ± 40% laterality).[16] Each gave written informed consent in compliance with the University of Memphis IRB.

2.2 Stimuli and task

We used a 7-token (hereafter "Tk1-Tk7") vowel continuum from /u/ to /ɑ/ synthesized in MATLAB (Natick, MA) via a conventional source-filter implementation. Each 100 ms token had a fundamental frequency of 100 Hz (i.e., male voice). 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 pilot data suggested vowels were more prone to nonlinear perceptual effects (see supplementary material for details in Fig. S1[17]). We delivered stimuli binaurally through insert earphones (ER-2; Etymotic Research, Elk Grove Village, IL) at 76 dB_A SPL. Sounds were controlled by MATLAB coupled to a TDT RP2 signal processor (Tucker-Davis Technologies, Alachua, FL).

There were three conditions based on how tokens were sequenced: (1) random presentation, and two sequential orderings presented serially between continuum end points and F1 frequencies, (2) forward /u/ to /ɑ/, 430–730 Hz (i.e., vowel lowering), and (3) backward /ɑ/ to /u/, 730–430 Hz (i.e., vowel raising). Forward and backward directions on such a continuum were expected to produce perceptual warpings (i.e., hysteresis).[10] Random and serial order conditions were presented in three different blocks (1 random, 1 forward, 1 backward), randomized between participants. We allowed breaks between blocks to avoid fatigue.

Within each condition, listeners heard 100 presentations of each vowel (total = 700 per block). On each trial, listeners rapidly reported which phoneme they heard with a binary keyboard response ("u" or "a"). Following their response, the interstimulus interval was jittered randomly between 800 and 1000 ms (20 ms steps, uniform distribution).

2.3 Behavioral data analysis

2.3.1 Psychometric function analysis

Identification scores were fit with 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.[14,18] Leftward/rightward shifts in $\beta 0$ location for the sequential versus random stimulus orderings would reveal changes in the perceptual boundary characteristic of perceptual nonlinearity.[10] These metrics were analyzed using a one-way mixed-model analysis of variance (ANOVA) (subjects = random factor) with a fixed effect of condition (three levels: random, forward, and backward) and Tukey–Kramer adjustments for multiple comparisons. Reaction times (RTs) were computed as the median response latency for each token per condition. RTs outside of 250–2000 ms were considered outliers (i.e., guesses or attentional lapses) and were excluded from analysis [n = 2487 trials (~7%) across all tokens/conditions/listeners].[13,14] RTs were analyzed using a two-way, mixed model ANOVA (subjects = random) with fixed effects of condition (three levels: random, forward, and backward) and token (seven levels).

2.3.2 Cross-classification analysis of behavioral response sequences

To determine the effect of sequential presentation order (i.e., forward versus backward F1) on behavioral responses, we performed cross-classification analysis on single-runs of the identification data (i.e., responses from tokens 1–7 or 7–1) in the Generalized Sequential Querier program.[19] This compared listeners' category labels for each continuum token (e.g., instances where Tk 3 presentations were labeled as "u" versus "a") when the stimulus continuum was presented in the forward (i.e., rising F1) versus backward (i.e., falling F1) direction. Biasing due to presentation order was quantified using Yule's Q, an index of standardized effect size transformed from an odds ratio; it varies from –1 to 1, which is superior to the odds ratio because it is relatively unskewed, affording more direct statistical analysis.[20] In the current application, a Q of +1 means "u" selected more in the forward F1 condition and "a" selected more in the backward F1 condition; a Q of −1 indicates the opposite pattern; and values effectively equal to 0 indicate presentation order had no effect on response selection. This analysis allowed us to determine whether the direction of stimulus presentation (i.e., increasing/decreasing F1) shifted listeners' category labels towards one end point of the continuum or the other (i.e., evidence of perceptual hysteresis). The non-0 responses at Tk3/Tk5 were used to classify participants as "hysteresis" versus "enhanced contrast" listeners (i.e., those showing late versus early biasing in their category labeling). See supplementary material for details on the cross-classification analysis results in Table S1.[17]

2.4 EEG recording procedures and analysis

2.4.1 EEG recording

Continuous EEGs were recorded during the speech identification task from 64 sintered Ag/AgCl electrodes at standard 10–10 scalp locations (NeuroScan Quik-Cap array).[21] Continuous data were sampled at 500 Hz (SynAmps RT amplifiers; Compumedics NeuroScan; Charlotte, NC) with an online passband of DC-200 Hz. Electrodes placed on the outer canthi of the eyes and superior/inferior orbit monitored ocular movements. Contact impedances were <10 kΩ. During acquisition, electrodes were referenced to an additional sensor placed ∼1 cm posterior to the Cz channel. Data were common average referenced for analysis.

2.4.2 Cluster-based permutation analysis

To reduce data dimensionality, channel clusters were computed by averaging adjacent electrodes over 5 *a priori* left/right frontocentral scalp areas as defined in previous speech ERP studies (see Fig. 2).[14,22] We used cluster-based permutation statistics[23] implemented in BESA® Statistics 2.1 (BESA, GmbH) to determine whether channel cluster ERP amplitudes differed with presentation order. This ran an initial *F*-test across the whole waveform (i.e., –200–800 ms), contrasting random, forward, and backward F1 conditions. This step identified time samples and channel clusters where neural activity differed between conditions ($p < 0.05$). Critically, BESA corrects for multiple comparisons across space and time. This was then followed by a second level analysis using permutation testing ($N = 1000$ resamples) to identify significant *post hoc* differences between pairwise stimulus conditions (i.e., random/forward/backward stimulus orderings). Contrasts were corrected with Scheffe's test using Bonferroni–Holm adjustments. Last, we repeated this analysis for tokens 3–5, representing stimuli surrounding the categorical boundary where warping was expected.

2.4.3 Distributed source analysis

We used Classical LORETA Analysis Recursively Applied (CLARA) distributed imaging with a 4 shell ellipsoidal head model [conductivities of 0.33 (brain), 0.33 (scalp), 0.0042 (bone), and 1.00 (cerebrospinal fluid)] on difference waves to determine the intracerebral sources that account for perceptual non-linearities in speech categorization.[24] Difference waves were computed as the voltage difference in ERPs for each of the three pairwise stimulus contrasts (i.e., random–forward; random–backward; forward–backward). All 64 electrodes were used (rather than the channel cluster subset) since full head coverage is needed to reconstruct inverse solutions. Source images were computed at a latency of 320 ms, where the scalp ERPs maximally differentiated stimulus order based on the cluster-based statistics [see Fig. 3(A)]. Correlations between changes in $\beta 0$ and CLARA activations evaluated which source regions predicted listeners' perceptual warping of speech categories.

3. Results

3.1 Behavioral data

3.1.2 Psychometric function data

Listeners perceived vowels categorically in all three presentation orderings as evidenced by their sigmoidal identification functions [Fig. 1(A)]. Slopes varied with presentation order ($F_{2,28} = 6.96$, $p = 0.0463$); this was driven by the forward condition producing stronger categorization than random ($p = 0.0364$) [Fig. 1(C)]. The categorical boundary did not appear to change with condition when analyzed at the group level ($F_{2,28} = 1.78$, $p = 0.1875$) [Fig. 1(D)].

RTs varied with presentation order ($F_{2,292} = 8.45$, $p = 0.0003$) and token ($F_{6,292} = 10.85$, $p < 0.0001$) [Fig. 1(B)]. Participants' labeling was slower for random compared to forward ($p = 0.0002$) and backward ($p = 0.0419$) presentation orders. RTs were also slower near the continuum midpoint versus end points (Tk4 versus Tk1/7: $p < 0.0001$), consistent with previous studies demonstrating slower RTs for category-ambiguous speech sounds.[7,14,25,26] Pooling orders, comparisons between the left/right sides of the continuum (Tk1,2,3 versus Tk5,6,7) indicated listeners responded to "ɑ" vowels faster than "u" vowels ($p < 0.0001$). This suggests sequential presentation of the continua, regardless of direction, improved speech categorization speeds.

Despite limited changes in boundary location at the group level [Fig. 1(D)], perceptual nonlinearities were subject to stark individual differences [Figs. 1(E)–1(G)]. Some listeners were consistent in their percept of individual tokens regardless of presentation order (i.e., "critical boundary" response pattern) ($n = 1$); others persisted with responses well beyond the putative category boundary at continuum midpoint (i.e., hysteresis) ($n = 9$); and other listeners changed responses earlier than expected (i.e., enhanced contrast) ($n = 4$). Response patterns were, however, highly stable *within* listener; a split-half analysis showed $\beta 0$ locations were strongly correlated between the first and last half of task trials ($r = 0.86$, $p < 0.0001$). This suggests that while perceptual nonlinearities (i.e., $\beta 0$ shifts) varied across listeners, response patterns were highly repeatable within individuals.

We performed further cross-classification analysis to characterize these individual differences in categorization nonlinearities. Table S1 shows participants' Yule's Q values Tk3/5 (i.e., tokens flanking the $\beta 0$), and, thus, their predominant "mode" of hearing the speech continua (see supplementary material for details on individual listening strategies[17]).

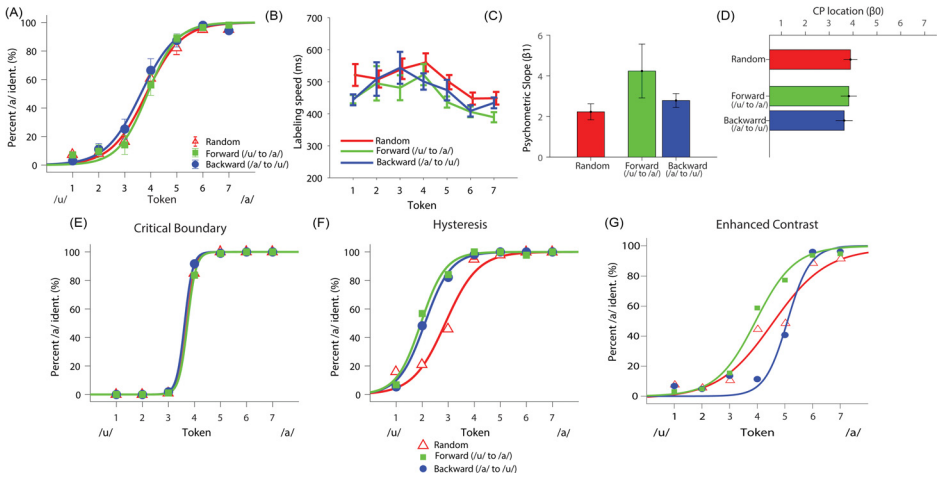

Fig. 1. Behavioral speech categorization is modulated by stimulus presentation order revealing nonlinearities in perception. (A) Perceptual psychometric functions for phoneme identification when continuum tokens are presented in random versus serial (forward: /u/→/ɑ/ versus backward: /ɑ/→/u/) order. (B) Reaction times for speech identification. Sequential presentation (i.e., forward and backward) led to faster speech labeling speeds than random presentation. (C) Psychometric function slope was steeper for forward compared to random presentation. (D) Boundary location did not vary at the group level (cf. individual differences; E–G). Individual differences reveal unique forms of perceptual nonlinearity across sub-classes of listeners (n = 3 representative subjects, one for each response pattern: s4, s2, s1). (E) Critical boundary listener (s4), where the individual selects the same response, regardless of presentation order. (F) Hysteresis listener (s1), where the prior percept continues beyond the expected perceptual boundary (midpoint) as measured in sequential presentation (cf. panel E). (G) Enhanced contrast listener (s2), where the category response flips earlier than expected during sequential presentation. See supplementary material for classifications of these listeners in Table S1(Ref. 17). Error bars = ±1 SEM.

Individuals with negative Qs showed hysteresis response patterns ($n = 9$), while those with positive Qs showed enhanced contrast patterns in perception ($n = 4$). Still others ($n = 2$) did not show perceptual nonlinearities and demonstrated neither hysteresis nor enhanced contrast.

3.2 Electrophysiological data

Figure 2 shows scalp ERP channel clusters to token 4 (critical stimulus at the perceptual boundary) across presentation orders (see supplementary material for raw ERP data in Fig. S2[17]). Cluster based permutation tests[23] also revealed nonlinear (stimulus order) effects emerging ~320 ms after speech onset, localized to left temporal areas of the scalp (omnibus ANOVA; $p = 0.03$) [Fig. 3(A), shading]. Condition effects were not observed in other channel clusters. *Post hoc* contrasts revealed order effects were driven by larger neural responses for the random versus forward F1 condition ($p = 0.003$). CLARA source reconstruction localized this nonlinear effect (i.e., $ERP_{random@Tk4} > ERP_{forward@Tk4}$) to underlying brain regions in bilateral superior temporal gyri (STG) and medial (MFG) and inferior (IFG) frontal gyri [Fig. 3(B)]. No differences were found when grouping neural responses by behavioral response patterns, including when accounting for differences in the listeners' categorical boundary. However, this might be expected given the low sample size ("n") within each subgroup.

We assessed the behavioral relevance of these neural findings via correlations between regional source activations (i.e., CLARA amplitudes at 320 ms) [Figs. 3(C) and 3(D)] and listeners' behavioral CP boundary locations ($\beta 0$). We found modulations in right MFG (rMFG) and left IFG with stimulus order were associated with behavioral $\beta 0$ shifts characteristic of perceptual warping but in opposite directions. Listeners with increased rMFG activation from random versus ordered (forward) stimulus presentation showed lesser movement of their perceptual boundary [Pearson's $r = -0.72$, $p = 0.0027$]. In

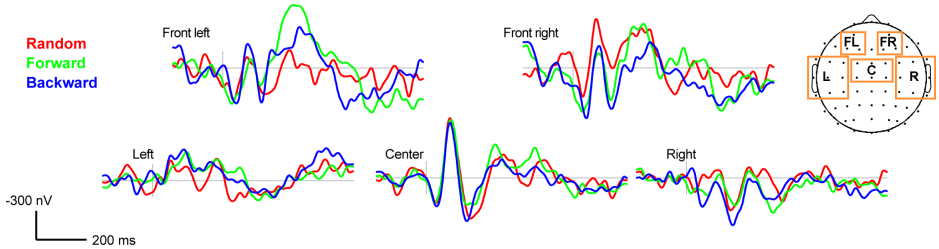

Fig. 2. Grand average ERPs (at Tk4 = critical boundary stimulus) for forward, backward, and random presentation order of the vowel continuum. Boxes = channel electrode cluster locations. Negative voltage plotted up.

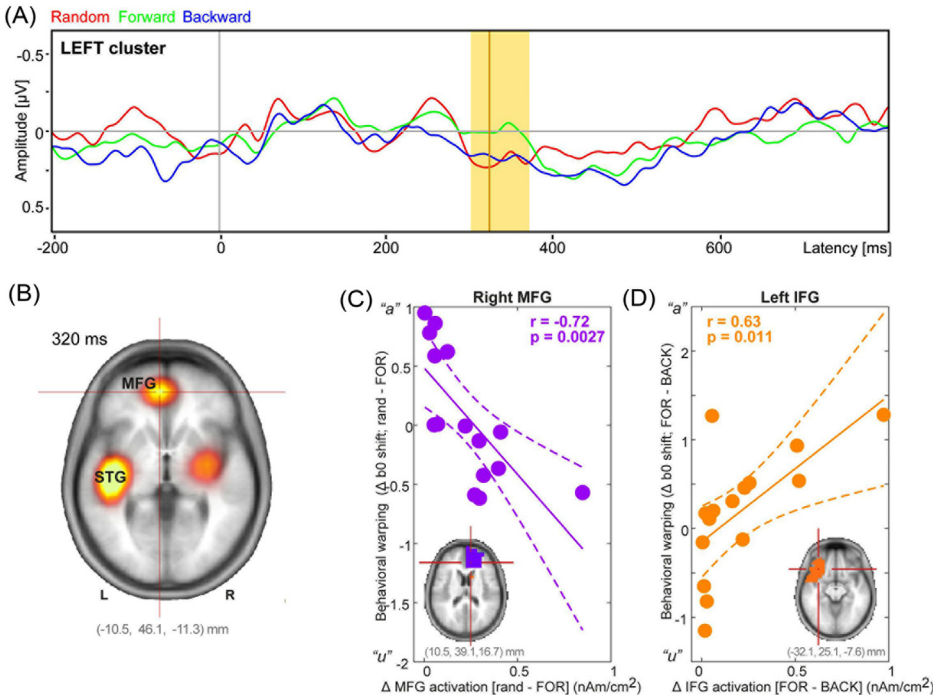

Fig. 3. Perceptual nonlinearities in the auditory cortical ERPs emerge by ~320 ms via interplay between frontotemporal cortices. (A) Cluster based permutation statistics contrasting responses to the identical Tk4 (continuum's midpoint) in random, backward, and forward conditions. Nonlinearities in speech coding emerge by ~300 ms (highlighted region) in the left channel cluster. Line = maximal difference (322 ms). Negative = up. (B) CLARA source imaging contrasting the difference in activations to Tk4 during random versus forward conditions. Nonlinearities in perceptual processing localize to bilateral superior temporal gyri and medial/inferior frontal gyri. (C) and (D) Brain–behavior correlations between the change in regional source activations and magnitude of hysteresis effect. Changes in right rMFG contrasting "randomness" (i.e., Δrandom–forward) are negatively associated with shifts in the CP boundary. Contrastively, modulations in left IFG contrast the *direction* of serial ordering (i.e., Δforward–backward) and are positively related to behavior.

contrast, those with increased left IFG activation contrasting stimulus direction (i.e., Δ forward versus backward) showed larger movement in $\beta 0$ location [$r = 0.63$, $p = 0.011$]. STG activations did not correlate with behavior (corrected p's > 0.05).

4. Discussion

By measuring EEG to acoustic–phonetic continua presented in different contexts (random, serial orderings), our data expose the brain mechanisms by which listeners assign otherwise identical speech tokens to categories depending on context. Behaviorally, perceptual nonlinearities were more prominent for vowels compared to CVs (see supplementary material[17]) and were subject to stark individual differences. Behavioral warping corresponded with neural effects emerging ~300 ms over left hemisphere with underlying sources in a frontotemporal circuit (bilateral STG, right MFG, left IFG). Our findings reveal stimulus presentation order strongly influences the neural encoding of phonemes and suggest that sequential warpings in speech perception emerge from top-down, dynamic modulation of early auditory cortical activity via frontal brain regions.

4.1 Perceptual nonlinearities in categorization are stronger for vowels than CVs

We found vowels elicited stronger perceptual warping (i.e., changes in the CP boundary) than CV tokens (see supplementary material[17]). Vowels are generally perceived less categorically than CVs.[1,12,27,28] With the vowel state space already being more flexible than consonants, listeners are more free to alter perception based on history of other vowels. Formant frequencies intrinsic to vowels are relatively continuous in their variations, but also static. In contrast, formant transitions in CVs allow frequency comparisons within the stimulus itself.[29,30] Vowel percepts are thus more ambiguous categorically, and consequently, more susceptible to contextual influences and individual differences.[31] Indeed, we find the magnitude and direction of perceptual warping strongly varies across listeners, consistent with prior work on perceptual hysteresis in both the auditory and visual domains.[10,32]

4.2 Perceptual warping of categories is subject to stark individual differences

Behaviorally, we found minimal group-level differences in psychometric functions, with only an increase in slope when in the forward /u/ to /ɑ/ direction versus random presentation. A change in identification slope indicates sequential

presentation led to more abrupt category changes. The reason behind this direction-dependent effect is unclear but could be related to differences in perceptual salience between continuum end points. We can rule out differences due to vowel loudness as both /u/ and /ɑ/ end points had nearly identical loudness according to ANSI (2007)[33] (/ɑ/= 71.9 phon; /u/ = 71.2 phon).[34] Alternatively, /ɑ/ might have been heard as being a more prototypical vowel (i.e., perceptual magnet),[35] perhaps owing to its higher frequency of occurrence in the English language.[36,37] Another explanation is that in the forward ordering, tokens were increasing in F1 frequency and previous work has demonstrated listeners are more sensitive to changes in rising versus falling pitch.[38,39] Thus, the increase in F1 may be more salient from a pitch (or spectral percept) standpoint. Conversely, RTs were faster in sequential compared to random presentation orders. RTs demonstrate the speed of processing, which increases (i.e., slows down) for more ambiguous or degraded tokens[7,30] and decreases (i.e., speeds up) for more prototypical tokens.[25] Faster RTs during sequential presentation suggest a quasi-priming effect whereby responses to adjacent tokens were facilitated by the preceding (phonetically similar) stimulus.

Behavioral changes in category boundary location were most evident at the individual rather than group level (cf. Refs. 8 and 40) and when speech tokens were presented sequentially. These findings suggest stimulus history plays a critical role in the current percept of phonemes. Listeners demonstrated three distinct response patterns (see supplementary material for hysteresis, enhanced contrast, and critical boundary shown in Table S1[17]) differences which were largely obscured at the group level. This is consistent with previous work demonstrating trial-by-trial differences in nonlinear dynamics of speech categorization.[9–11] Critically, response patterns were highly stable *within* individuals, suggesting listeners have a dominant response pattern and/or apply different decision strategies (cf. biases) during categorization. This latter interpretation is also supported by the different regional activation patterns and their behavioral correlations. It is also reminiscent of lax versus strict observer models in signal detection frameworks where, for suprathreshold stimuli, listeners' response selection is primarily determined by their internal bias (i.e., preference for tokens at one end of the continuum).[41]

4.3 Electrophysiological correlates of perceptual warping

ERPs revealed late (~320 ms post-stimulus) differences in response to token 4 (i.e., categorical boundary) between forward and random conditions over the left hemisphere. Sound-evoked responses in auditory cortex typically subside after ~250 ms.[42,43] This suggests the stimulus order effects observed in our speech ERPs likely occur in higher-order brain regions subserving linguistic and/or attentional processing. The leftward lateralization of responses also suggests context-dependent coding might be mediated by canonical language-processing regions (e.g., Broca's area).[44] Indeed, source analysis confirmed engagement of extra-auditory brain areas including IFG and MFG whose activations scaled with listeners' perceptual shifts in category boundary. In contrast, auditory STG, though active during perceptual warping, did not correlate with behavior, *per se*.

Beyond its established role in speech-language processing, left IFG is heavily involved in category decisions, particularly under states of stimulus uncertainty (i.e., randomness, noise).[7,14,31] Related, we find *direction-related* modulations in the perceptual warping of speech categories (to otherwise identical sounds) are predicted by left IFG engagement. IFG involvement in our tasks is consistent with notions that frontal brain regions help shape behavioral category-level predictions at the individual level.[45] Contrastively, rMFG correlated with changes in behavior between random versus forward stimulus presentation, a contrast of ordered versus unordered sequencing. MFG regulates behavioral reorienting and serves to break (i.e., gate) attention during sensory processing.[46] Additionally, it is active when holding information in working memory, such as performing mental calculations,[47] and has been implicated in processing ordered numerical sequences and counting.[48] The observed perceptual nonlinearities induced by serial presentation might therefore be driven by such buffer and comparator functions of rMFG as listeners hold prior speech sounds in memory and compare present to previous sensory–memory traces. In contrast, un-ordered speech presented back-to-back would not load those operations and thus, may explain the reduced rMFG activity for random presentation. The simultaneous activation of canonical auditory areas (STG) concurrent with these two frontal regions leads us to infer that while auditory cortex is sensitive to category structure (present study; Refs. 7 and 14) top-down modulations from frontal lobes dynamically shapes category percepts online during speech perception.

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.

References and links

[1] D. B. Pisoni, "Auditory and phonetic memory codes in the discrimination of consonants and vowels," Percept. Psychophys. **13**, 253–260 (1973).
[2] A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy, "Perception of the speech code," Psychol. Rev. **74**, 431–461 (1967).
[3] P. K. Kuhl, "Theoretical contributions of tests on animals to the special-mechanisms debate in speech," Exp. Biol. **45**, 233–265 (1986).
[4] G. M. Bidelman, C. Pearson, and A. Harrison, "Lexical influences on categorical speech perception are driven by a temporoparietal circuit," bioRxiv:2020.08.11.246793 (2020).

[5]W. F. Ganong III and R. J. Zatorre, "Measuring phoneme boundaries four ways," J. Acoust. Soc. Am. **68**, 431–439 (1980).
[6]K. Mankel and M. Bidelman, "Inherent auditory skills rather than formal music training shape the neural encoding of speech," Proc. Natl. Acad. Sci. U. S. A. **115**, 13129–13134 (2018).
[7]J. A. Carter and G. M. Bidelman, "Auditory cortex is susceptible to lexical influence as revealed by informational vs energetic masking of speech categorization," Brain Res. **1759**, 147385 (2021).
[8]R. L. Diehl, J. L. Elman, and S. B. McCusker, "Contrast effects on stop consonant identification," J. Exp. Psychol. Hum. Percept. Perform. **4**, 599 (1978).
[9]N. Nguyen, S. Wauquier, and B. Tuller, "The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories," in *Approaches to Phonological Complexity* (De Gruyter Mouton, Berlin/Boston, 2009), pp. 191–218.
[10]B. Tuller, P. Case, M. Ding, and J. A. S. Kelso, "The nonlinear dynamics of speech categorization," J. Exp. Psychol. Hum. Percept. Perform. **20**, 3–16 (1994).
[11]B. Tuller, M. G. Jantzen, and V. K. Jirsa, "A dynamical approach to speech categorization: Two routes to learning," New Ideas Psychol. **26**, 208–226 (2008).
[12]M. Studdert-Kennedy, A. M. Liberman, K. S. Harris, and F. S. Cooper, "Motor theory of speech perception: A reply to Lane's critical review," Psychol. Rev. **77**(3), 234–249 (1970).
[13]G. M. Bidelman, S. Moreno, and C. Alain, "Tracing the emergence of categorical speech perception in the human auditory system," NeuroImage **79**, 201–212 (2013).
[14]G. M. Bidelman and B. Walker, "Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network," NeuroImage **201**, 116022 (2019).
[15]E. Liebenthal, R. Desai, M. Ellingson, B. Ramachandran, A. Desai, and J. Binder, "Specialization along the left superior temporal sulcus for auditory categorization," Cereb. Cortex **20**, 2958–2970 (2010).
[16]R. C. Oldfield, "The assessment and analysis of handedness: The Edinburgh inventory," Neuropsychologia **9**, 97–113 (1971).
[17]See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0009896 for details.
[18]G. M. Bidelman, M. W. Weiss, S. Moreno, and C. Alain, "Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians," Eur. J. Neurosci. **40**, 2662–2673 (2014).
[19]Information on v 5.1.23 is available at https://www.mangold-international.com/en/products/software/gseq
[20]R. Bakeman and V. Quera, *Sequential Analysis and Observational Methods for the Behavioral Sciences* (Cambridge University Press, Cambridge, 2011)
[21]R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," Clin. Neurophysiol. **112**, 713–719 (2001).
[22]C. Marie and L. J. Trainor, "Development of simultaneous pitch encoding: Infants show a high voice superiority effect," Cereb. Cortex **23**, 660–669 (2013).
[23]E. Maris and R. Oostenveld, "Nonparametric statistical testing of EEG-and MEG-data," J. Neurosci. Methods **164**, 177–190 (2007).
[24]M. Scherg, P. Berg, N. Nakasato, and S. Beniczky, "Taking the EEG back into the brain: The power of multiple discrete sources," Front. Neurol. **10**, 855 (2019).
[25]D. B. Pisoni and J. Tash, "Reaction times to comparisons within and across phonetic categories," Percept. Psychophys. **15**, 285–290 (1974).
[26]R. Reetzke, Xie, Z., Llanos, F., and B. Chandrasekaran, "Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood," Curr. Biol. **28**, 1419–1427 (2018).
[27]D. B. Pisoni, "Auditory short-term memory and vowel perception," Mem. Cognit. **3**, 7–18 (1975).
[28]C. F. Altmann, M. Uesaki, K. Ono, M. Matsuhashi, T. Mima, and H. Fukuyama, "Categorical speech perception during active discrimination of consonants and vowels," Neuropsychologia **64**, 13–23 (2014).
[29]Y. Xu, J. T. Gandour, and A. L. Francis, "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," J. Acoust. Soc. Am. **120**, 1063–1074 (2006).
[30]G. M. Bidelman, L. C. Bush, and A. M. Boudreaux, "Effects of noise on the behavioral and neural categorization of speech," Front. Neurosci. **14**, 1–13 (2020).
[31]G. M. Bidelman, C. Pearson, and A. Harrison, "Lexical influences on categorical speech perception are driven by a temporoparietal circuit," J. Cogn. Neurosci. **33**, 840–852 (2021).
[32]A. Sayal, A. T Sousa, J. V. Duarte, G. N. Costa, R. A. Martins, and M. Castelo-Branco, "Identification of competing neural mechanisms underlying positive and negative perceptual hysteresis in the human visual system," NeuroImage **221**, 117153 (2020).
[33]ANSI S3.4-2007.
[34]B. C. J. Moore, B. R. Glasberg, and T. Baer, "A model for the prediction of thresholds loudness and partial loudness," J. Audio Eng. Soc. **45**, 224–240 (1997).
[35]P. Iverson and P. K. Kuhl, "Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism?," Percept. Psychophys **62**, 874–886 (2000).
[36]R. E. Hayden, "The relative frequency of phonemes in General-American English," Word **6**, 217–223 (1950).
[37]M. A. Mines, B. F. Hanson, and J. E. Shoup, "Frequency of occurrence of phonemes in conversational English," Lang. Speech **21**, 221–241 (1978).
[38]M. E. H. Schouten, "Identification and discrimination of sweep tones," Percept. Psychophys. **37**, 369–376 (1985).
[39]H. Luo, A. Boemio, M. Gordon, and D. Poeppel, "The perception of FM sweeps by Chinese and English listeners," Hear. Res. **224**, 75–83 (2007).
[40]A. F. Healy and B. H. Repp, "Context independence and phonetic mediation in categorical perception," J. Exp. Psychol. Hum. Percept. Perform. **8**, 68 (1982).
[41]D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966), Vol. **xi**, p. 455.
[42]K. E. Crowley and I. M. Colrain, "A review of the evidence for P2 being an independent component process: Age, sleep and modality," Clin. Neurophysiol. **115**, 732–744 (2004).

[43]A. P. Fonaryova, G. O. Dove, and M. J. Maguire, "Linking brainwaves to the brain: An ERP primer," Dev. Neuropsychol. **27**, 183–215 (2005).
[44]G. Hickok, M. Costanzo, R. Capasso, and G. Miceli, "The role of Broca's area in speech perception: Evidence from aphasia revisited," Brain Lang. **119**, 214–220 (2011).
[45]P. Fuhrmeister and E. B. Myers, "Structural neural correlates of individual differences in categorical perception," Brain Lang. **215**, 104919 (2021).
[46]S. Japee, K. Holiday, M. D. Satyshur, I. Mukai, and L. G. Ungerleider, "A role of right middle frontal gyrus in reorienting of attention: A case study," Front. Syst. Neurosci. **9**, 23 (2015).
[47]M. Arsalidou, M. Pawliw-Levac, M. Sadeghi, and J. Pascual-Leone, "Brain areas associated with numbers and calculations in children: Meta-analyses of fMRI studies," Dev. Cogn. Neurosci. **30**, 239–250 (2018).
[48]E. Zaleznik and J. Park, "The neural basis of counting sequences," Neuroimage **237**, 118146 (2021).

Pilot experiment: Characterizing perceptual warping for vowels vs. CVs

We first examined whether perceptual warping of categories varies among different speech sounds (vowels vs. consonants). To this end, we ran a pilot sample that included N=5 young adults. Participants were native speakers of American English and reported normal hearing. We used 7-step continua of vowels (/u/ to /ɑ/) with F1 frequencies spanning from 430 to 730 Hz and consonant-vowel (CV) syllables (/dɑ/ to /gɑ/) used in previous studies[7,44,45]. Listeners were instructed to listen to these stimuli through headphones and respond by clicking on an onscreen button whether they heard “oo” or “ah” in the vowel conditions and “da” or “ga” in the CV condition. With each condition, listeners heard 10 repetitions of each token (total = 70 tokens per condition). The pilot task was conducted via internet-based data collection using paradigms coded in E-Prime 3.0 delivered using E-Prime Go[46].

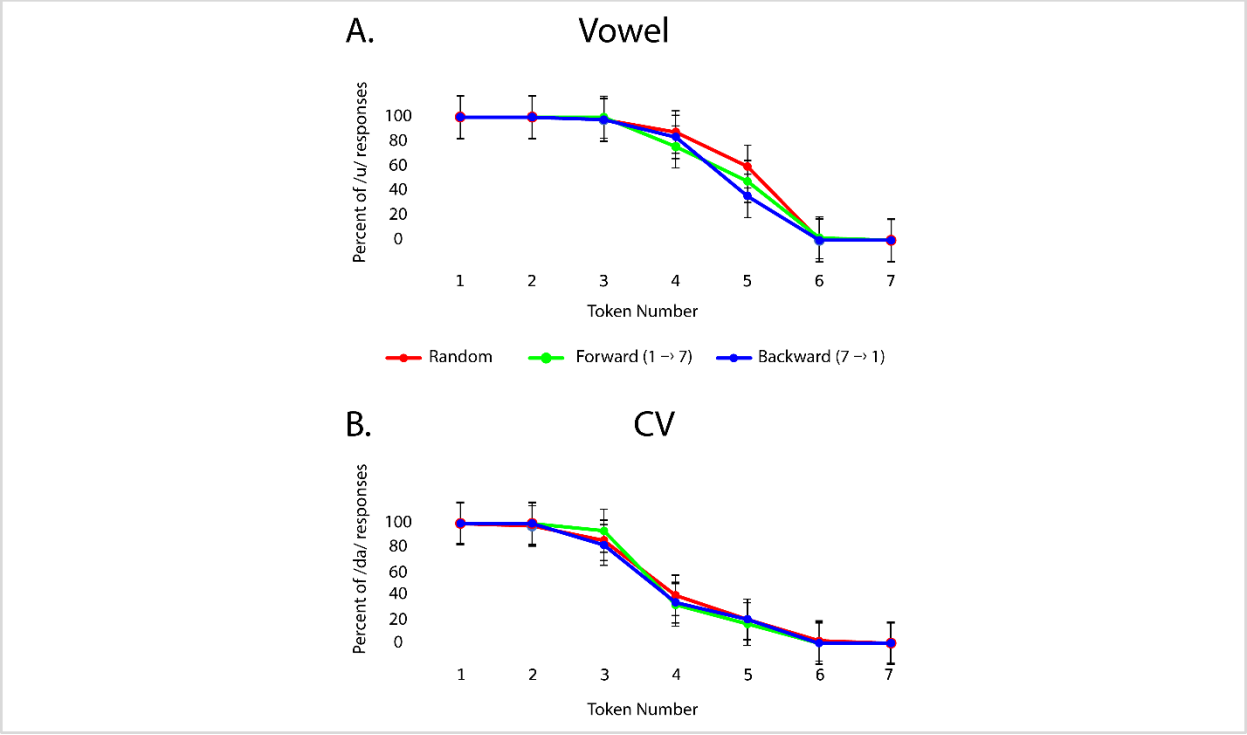

Figure S1: Psychometric functions (n=5), comparing the identification for **(A)** vowels and **(B)** consonant-vowel syllables (CVs). Vowels exhibited more nonlinear response patterns than CVs as evidenced by the more salient movement of the perceptual boundary (e.g., see Tk4-Tk5). Error bars = ±1 s.e.m.

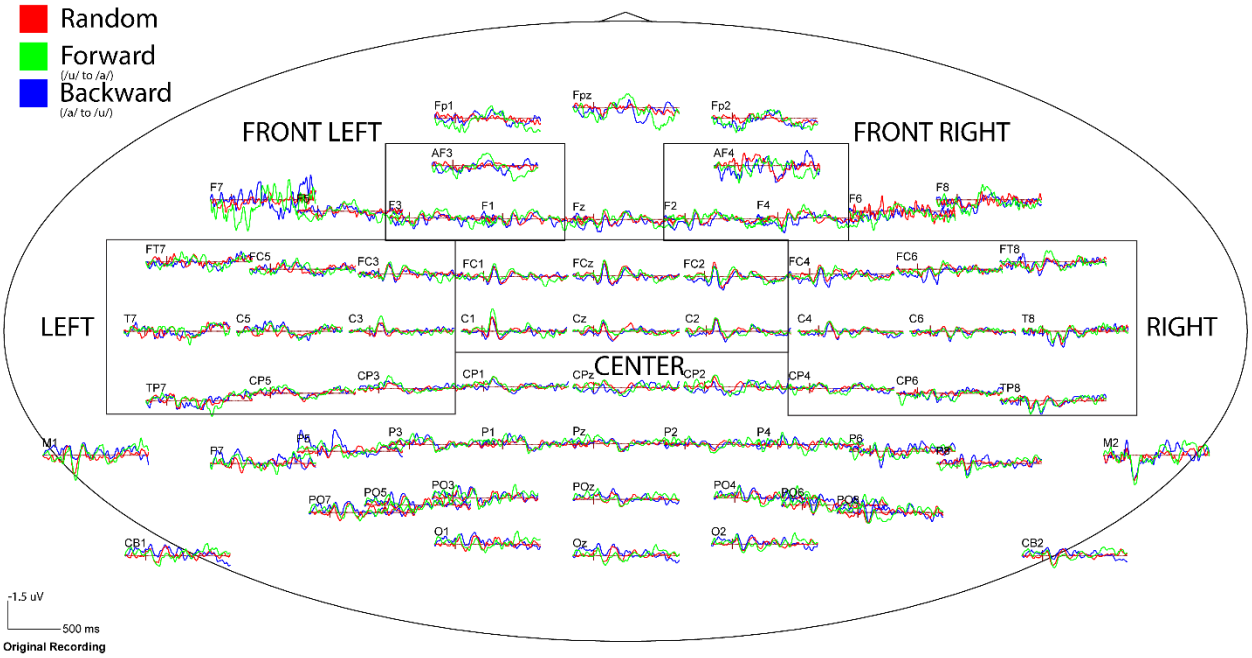

Figure S2: Scalp topography of ERPs. Boxes denote channel clusters used in the primary analysis. Negative is plotted up.

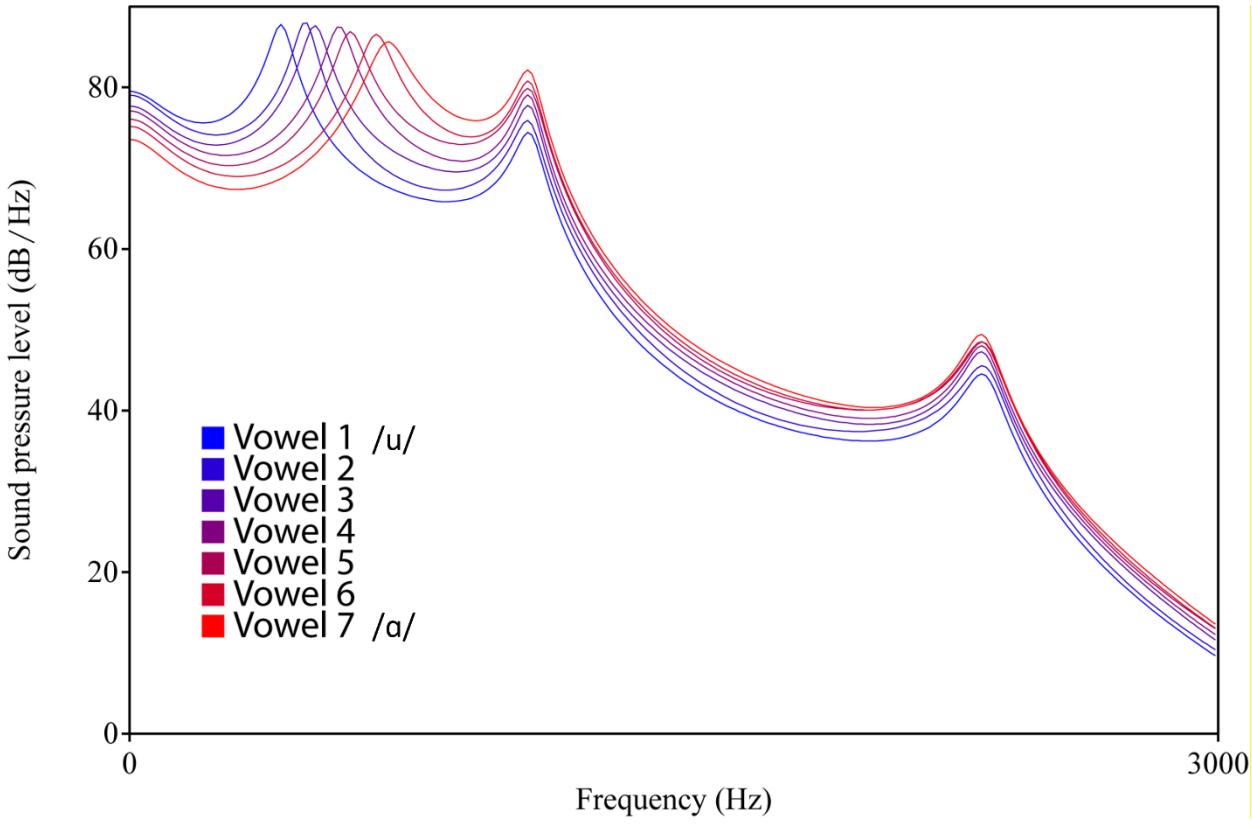

Figure S3: The formants of the tokens used in this project.

Subject Num	Tk3 Yule's Q	Tk5 Yule's Q	Response Pattern (listener type)
S1†	0.87*	0.75*	Hysteresis
S2†	-0.08	-0.66*	Enhanced Contrast
S3	0.00	-0.38*	Enhanced Contrast
S4†	0.00	0.00	Critical Boundary
S5	0.84*	0.49*	Hysteresis
S6	0.77*	0.00	Hysteresis
S7	1.00*	0.00	Hysteresis
S8	1.00*	0.74*	Hysteresis
S9	0.00	0.59*	Nil
S10	0.44*	0.00	Hysteresis
S11	0.76*	0.48*	Hysteresis
S12	0.23	-0.40*	Enhanced Contrast
S13	0.60*	0.30	Hysteresis
S14	0.00	-0.88*	Enhanced Contrast
S15	0.76*	0.17	Hysteresis

Table S1: Yule's Q values for Tk3/5 (i.e., tokens flanking the expected β0) and response patterns by participant. More positive/negative Yule's Q denotes hysteresis/enhanced contrast response patterns, respectively. *Yule's Q of medium-to-large effect size $|Q| \geq 0.33$. †Individuals shown in **Fig. 1E-G**