

**Title:** Nonlinear dynamics in auditory cortical activity reveal the neural basis of perceptual warping in speech categorization

**Author(s):** Jared A. Carter<sup>1,2</sup>, Eugene H. Buder<sup>1,2</sup> and Gavin M. Bidelman<sup>1,2,3</sup>

<sup>1</sup>Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA

<sup>2</sup>School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA

<sup>3</sup>University of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, TN, USA

**Author contributions:** J.A.C. & G.M.B. designed the experiment, J.A.C. collected the data, J.A.C., E.H.B., and G.M.B. analyzed the data and wrote the paper.

**Address for editorial correspondence:**

Jared A. Carter, M.S.

School of Communication Sciences & Disorders

University of Memphis

4055 North Park Loop

Memphis, TN, 38152

TEL: (682) 234-2839

EMAIL : [jcrter29@memphis.edu](mailto:jcrter29@memphis.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 nonlinearities. Behaviorally, sequential order shifted listeners' categorical boundary vs. 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 magnitudes. Findings demonstrate that top-down, stimulus history effects on speech categorization are governed by interactions between frontotemporal brain regions.

**Keywords:** auditory event-related potentials (ERPs); categorical perception (CP); hysteresis; cross-classification analysis; nonlinear speech processing

# 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, but reducing acoustic cues to discrete categories enables more efficient use of speech sounds for linguistic processing<sup>1,2</sup>. The extent to which phonetic speech categories built 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<sup>3</sup>. Alternatively, there is ample evidence that top-down processing influences speech categorization as suggested by enhancements observed in highly proficient listeners<sup>4-7</sup> and biasing effects, when individuals hear a different category depending on the surrounding speech context<sup>8</sup>.

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)<sup>9-11</sup>. Both stop consonant and vowel continua can produce context-dependent shifts in perception, though stronger perceptual warping occurs with more ambiguous speech sounds<sup>12</sup>.

Event-related brain potentials (ERPs) have been used to examine the neural underpinnings of speech categorization<sup>13-15</sup>. 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<sup>13</sup>. 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<sup>10</sup>. 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 vs. serial (forward or backward) order. Based on previous studies examining nonlinear dynamics<sup>9,10</sup> and top-down influences in

speech CP<sup>4,5,7</sup>, we hypothesized (1) the location of listeners' perceptual boundary would shift according to the direction of stimulus presentation and (2) perceptual warping would be accompanied by late modulations in the ERPs.

## 2. MATERIALS & METHODS

### 2.1 Participants

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

### 2.2 Stimuli & task

We used a 7-step vowel continuum from /u/ to /a/. 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 (/a/). We selected vowels over consonant-vowel (CV) syllables because pilot data suggested vowels were more prone to nonlinear perceptual effects (see **Figure S1**). We delivered stimuli binaurally through insert earphones (ER-3A) at 76 dB<sub>A</sub> 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 endpoints and F1 frequencies (2) forward /u/ to /a/, 430 Hz to 730 Hz, and (3) backward /a/ to /u/, 730 to 430 Hz). Forward and backward directions on such a continuum were expected to produce perceptual warpings (i.e., hysteresis)<sup>10</sup>. Random and serial order conditions were presented in different blocks (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<sup>14,17</sup>. 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<sup>10</sup>. These metrics were analyzed using a one-way mixed-model ANOVA (subjects = random factor) with a fixed effect of condition (3 levels: random F1, forward F1, backward F1) 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<sup>13,14</sup>. RTs were analyzed using a two-way, mixed model ANOVA (subjects = random) with fixed effects of condition (3 levels: random F1, forward F1, backward F1) and token (7 levels).

### 2.3.2 Cross-classification analysis of behavioral response sequences

To determine the effect of sequential presentation order (i.e., forward vs. 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 Quierier program (v 5.1.23; Mangold International <https://www.mangold-international.com/en/products/software/gseq>). This compared listeners' category labels to specific tokens (e.g., trials where token 3 was labeled as “u” vs. “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 to vary from -1 to 1 which is superior to the odds ratio in being relatively unskewed, affording more direct statistical analysis<sup>18</sup>. 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 endpoint 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” vs. “enhanced contrast” listeners (i.e., those showing late vs. early biasing in their category labeling). See Table S1 for details.

## 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)<sup>19</sup>. Continuous data were sampled at 500 Hz (SynAmps RT amplifiers; Compumedics NeuroScan) 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, “super-channel” clusters were computed by averaging adjacent electrodes over 5 left/right frontocentral scalp areas (see **Fig. 3**)<sup>14,20</sup>. We used cluster-based permutation statistics<sup>21</sup> implemented in BESA® Statistics 2.1 to determine whether super-channel 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 super-channels 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. Lastly, we repeated this analysis for tokens 3-5, representing stimuli surrounding the categorical boundary where warping was most 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 the difference wave to determine the intracerebral sources that account for perceptual non-linearities in speech categorization<sup>22</sup>. Source images were computed at a latency of 320 ms, where the scalp ERPs differentiated stimulus order based

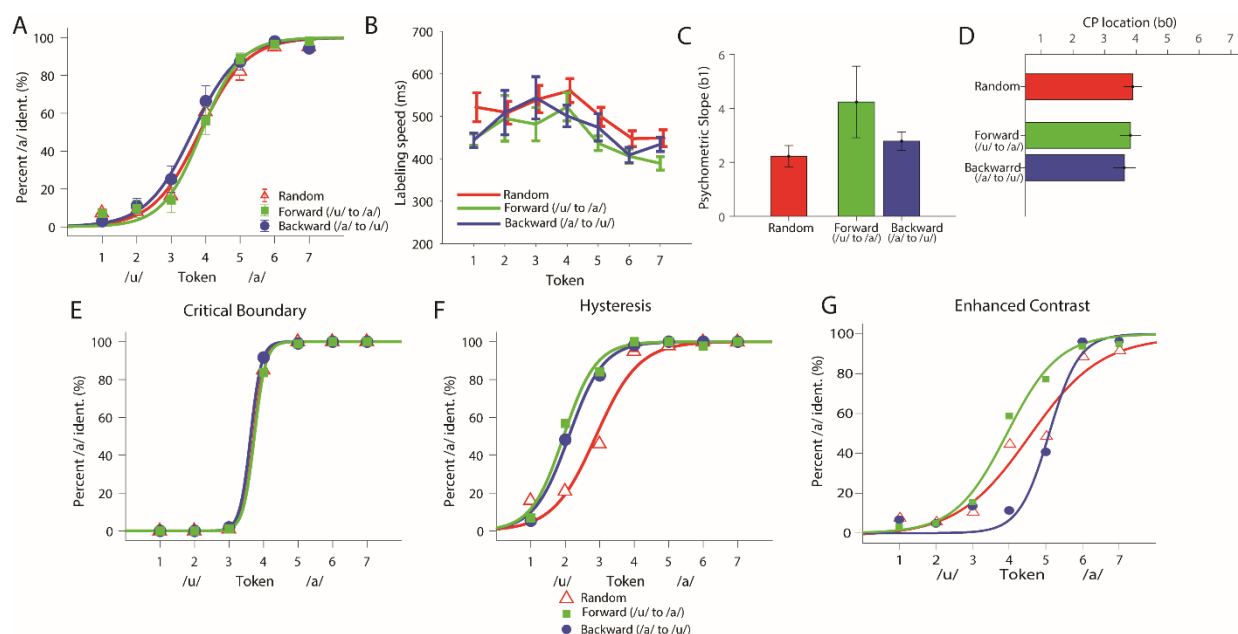
on the cluster-based statistics (see **Fig. 4**). Correlations between changes in  $\beta_0$  and CLARA activations evaluated which source regions predicted listeners' perceptual warping of speech categories.

### 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. 1A**). 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. 1C**). 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. 1D**).



**Figure 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 vs. serial (forward: /u/→/a/ vs. backward: /a/→/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; Fig. 2). Individual differences reveal

unique forms of perceptual nonlinearity across sub-classes of listeners ( $n=3$  representative subjects). **(E)** Critical boundary listener, where the individual selects the same response regardless of presentation order. **(F)** Hysteresis listener, where the prior percept continues beyond the expected perceptual boundary (midpoint) as measured in sequential presentation (cf. panel E). **(G)** Enhanced contrast listener, where the category response flips earlier than expected during sequential presentation. See also Table 1 for classifications of these listeners. Error bars =  $\pm 1$  s.e.m.

RTs varied with presentation order ( $F_{2,292} = 8.45$ ,  $p = 0.0003$ ) and token ( $F_{6,292} = 10.85$ ,  $p < 0.0001$ ) (**Fig. 1B**). Participants' labeling was slower for random compared to forward ( $p = 0.0002$ ) and backward ( $p = 0.0419$ ) presentation orders. RTs were also slower nearly the continuum midpoint vs. endpoints (Tk4 vs. Tk1/7:  $p < 0.0001$ ), consistent with previous studies demonstrating slower RTs for category-ambiguous speech sounds<sup>7,14,23,24</sup>. Pooling orders, comparisons between the left/right sides of the continuum (Tk1,2,3 vs. Tk5,6,7) indicated listeners responded to “a” 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. 1D**), perceptual nonlinearities were subject to stark individual differences (**Fig. 1E-G**). Some listeners were consistent in their percept of individual tokens regardless of presentation order (i.e., “critical boundary” response pattern); others persisted with responses well beyond the putative category boundary at continuum midpoint (i.e., hysteresis); and other listeners changed responses earlier than expected (i.e., enhanced contrast). 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.

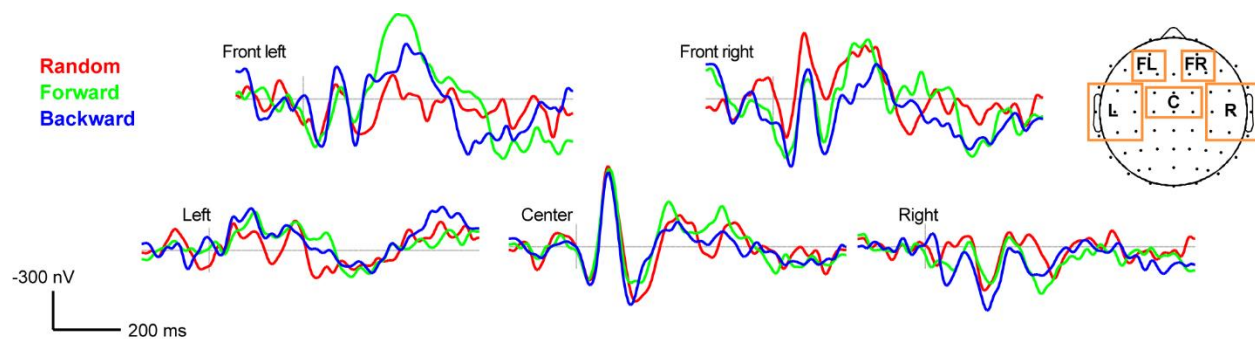
We performed further cross-classification analysis to characterize these individual differences in categorization nonlinearities. **Table S1** shows participants' Yule's Q values Tk3/5 (i.e., tokens flanking the  $\beta 0$ ), and, thus, their predominant “mode” of hearing the speech continua. Individuals 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

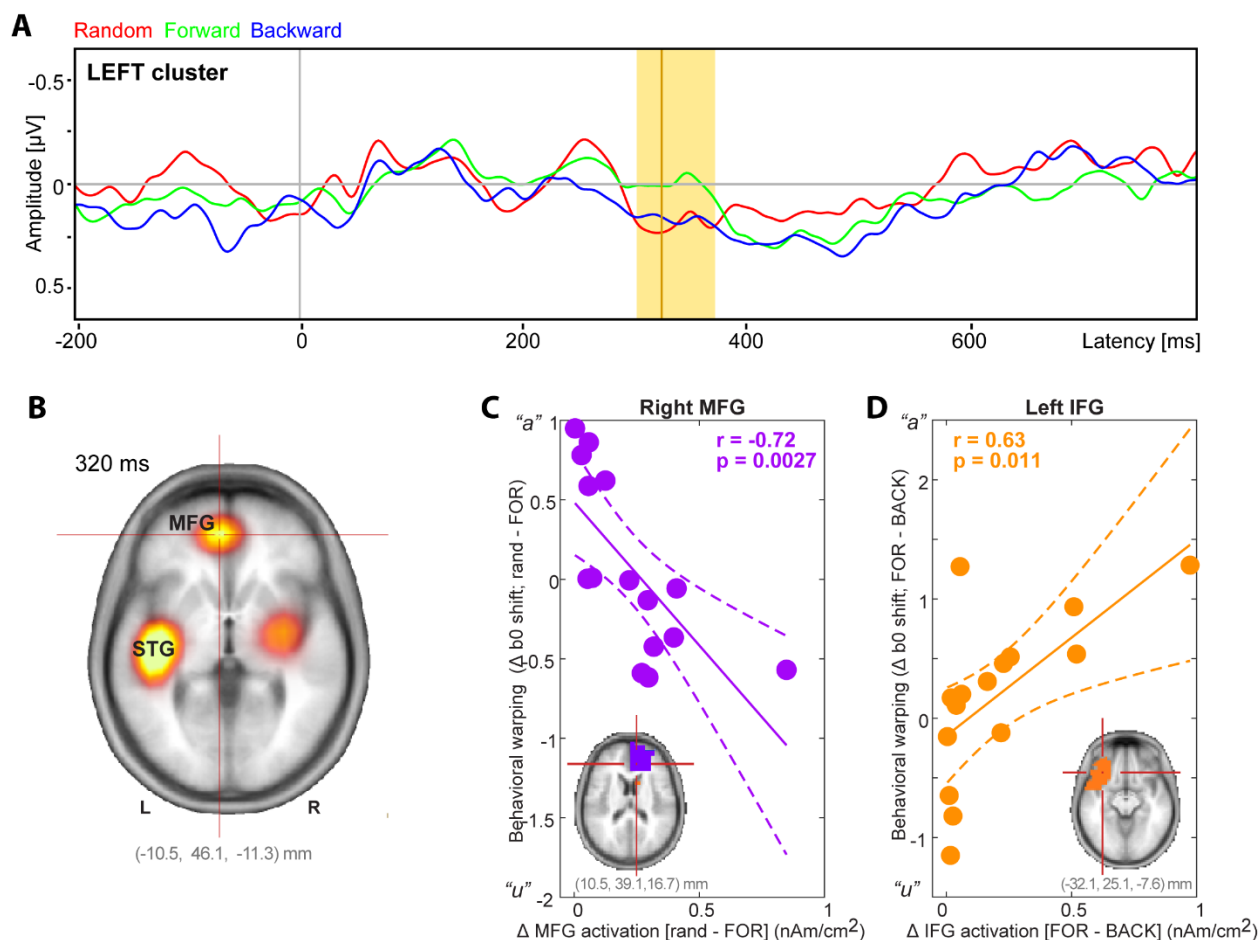
**Figure 2** shows scalp ERP “super channels” to token 4 (critical stimulus at the perceptual boundary) across presentation orders (see Fig. S2 for raw ERP data). Cluster based permutation tests<sup>21</sup> also revealed nonlinear (stimulus order) effects emerging ~320 ms after speech onset, localized to left temporal areas of the scalp (omnibus ANOVA;  $p=0.03$ ; **Fig. 3A**, shading). Post hoc contrasts revealed order effects were driven by larger neural responses for the random vs. forward F1 condition ( $p = 0.003$ ). CLARA source reconstruction localized this nonlinear effect (i.e.,  $ERP_{\text{random@Tk4}} > ERP_{\text{forward@Tk4}}$ ) to underlying brain regions in bilateral superior temporal gyri (STG) and medial (MFG) and inferior (IFG) frontal gyri (**Fig. 3B**).



**Figure 2:** Grand average ERPs (at Tk4 = critical boundary stimulus) for forward, backward, and random presentation order of the vowel continuum. Boxes = “super-channel” electrode clusters. Negative voltage plotted up.

We assessed the behavioral relevance of these neural findings via correlations between regional source activations (i.e., CLARA amplitudes @ 320 ms; **Fig. 3C-D**) and listeners’ behavioral CP boundary locations ( $\beta_0$ ). We found modulations in right MFG 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 vs. ordered (forward) stimulus presentation showed lesser movement of their perceptual boundary [Pearson’s  $r=-0.72$ ,  $p=0.0027$ ]. In contrast, those with increased left IFG activation contrasting stimulus

direction (i.e.,  $\Delta$  forward vs. backward) showed larger movement in  $\beta\theta$  location [ $r=0.63$ ,  $p=0.011$ ]. STG activations did not correlate with behavior (corrected  $p$ 's  $> 0.05$ ).



**Figure 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 (stimulus at the continuum's midpoint) in random, backward, and forward conditions. Nonlinearities in speech coding emerge by ~300 ms (highlighted region) in the left super-channel. Line = maximal difference (322 ms). Negative = up. (B) CLARA source imaging contrasting the difference in activations to Tk4 during random vs. forward conditions. Nonlinearities in perceptual processing localize to bilateral superior temporal gyri and medial/inferior frontal gyri. (C-D) Brain-behavior correlations between the change in regional source activations and magnitude of hysteresis effect. Changes in right rMFG contrasting "randomness" (i.e.,  $\Delta$  random - forward) are negatively

associated with shifts in the CP boundary. Contrastively, modulations in left IFG contrast the *direction* of serial ordering (i.e.,  $\Delta$ forward – backward) and are positively related to behavior.

## 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 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.

### *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. Vowels are generally perceived less categorically than CVs<sup>12,25-27</sup>. With the vowel state space already being more flexible than consonants, listeners are more free to alter perception based on prior 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<sup>28,29</sup>. Vowel percepts are thus more ambiguous categorically, and consequently more susceptible to contextual influences and individual differences<sup>30</sup>. 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<sup>10,31</sup>.

### *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 /a/ 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 endpoints. We can rule out differences due

to vowel loudness as both /u/ and /a/ endpoints had nearly identical loudness according to ANSI S3.4 (2007) (/a/ = 71.9 phon; /u/ = 71.2 phon)<sup>32</sup>. Alternatively, /a/ might have been heard as being a more prototypical vowel (i.e., perceptual magnet)<sup>33</sup>. 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<sup>7,29</sup> and decreases (i.e., speeds up) for more prototypical tokens<sup>23</sup>. 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.<sup>8,34</sup>) 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 (Table S1; hysteresis, enhanced contrast, critical boundary), 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<sup>9-11</sup>. 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 vs. 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)<sup>35</sup>.

### ***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<sup>36,37</sup>. 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)<sup>38</sup>. 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 differentially 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)<sup>7,30,39</sup>. Related, we find *direction-related* modulations in the perceptual warping of speech categories (to an otherwise identical sound) are predicted by left IFG engagement. This is consistent with notions that frontal brain regions help shape behavioral category-level predictions at the individual level<sup>40</sup>. Contrastively, rMFG correlated with changes in behavior between random vs. forward stimulus presentation, a contrast of ordered vs. unordered sequencing. MFG regulates behavioral reorienting and serves to break (i.e., gate) attention during sensory processing<sup>41</sup>. Additionally, it is active when holding information in working memory, such as performing mental calculations<sup>42</sup>, and has been implicated in processing ordered numerical sequences and counting<sup>43</sup>. 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 such 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; <sup>7,39</sup>), top-down modulations from frontal lobes dynamically shapes category percepts online during speech perception.

## Acknowledgements

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. [gmbdlman@memphis.edu].

# References

- <sup>1</sup>Pisoni, D.B., "Auditory and phonetic memory codes in the discrimination of consonants and vowels\*," *Perception & Psychophysics*, 13, 253-260 (1973).
- <sup>2</sup>Liberman, A.M., et al., "Perception of the Speech Code," *Psychological Review*, 74, 431-461 (1967).
- <sup>3</sup>Kuhl, P.K., "Theoretical contributions of tests on animals to the special-mechanisms debate in speech," *Experimental Biology*, 45, 233-265 (1986).
- <sup>4</sup>Bidelman, G.M., et al., "Lexical influences on categorical speech perception are driven by a temporoparietal circuit," *bioRxiv*, 2020.08.11.246793 (2020).
- <sup>5</sup>Ganong III, W.F. and R.J. Zatorre, "Measuring phoneme boundaries four ways," *The Journal of the Acoustical Society of America*, 68, 431-439 (1980).
- <sup>6</sup>Bidelman, G.M. and K. Mankel, "Inherent auditory skills rather than formal music training shape the neural encoding of speech," *Proceedings of the National Academy of Sciences of the United States of America*, 115, 13129-13134 (2018).
- <sup>7</sup>Carter, J.A. and G.M. Bidelman, "Auditory cortex is susceptible to lexical influence as revealed by informational vs. energetic masking of speech categorization," *Brain Research*, 1759, 147385 (2021).
- <sup>8</sup>Diehl, R.L., et al., "Contrast effects on stop consonant identification," *Journal of Experimental Psychology: Human Perception and Performance*, 4, 599 (1978).
- <sup>9</sup>Nguyen, N., et al., "The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories", in *Approaches to phonological complexity*. 2009, De Gruyter Mouton. p. 191-218.
- <sup>10</sup>Tuller, B., et al., "The Nonlinear Dynamics of Speech Categorization," *Journal of Experimental Psychology: Human Perception and Performance*, 20, 3-16 (1994).
- <sup>11</sup>Tuller, B., et al., "A dynamical approach to speech categorization: Two routes to learning," *New Ideas in Psychology*, 26, 208-226 (2008).
- <sup>12</sup>Studdert-Kennedy, M., et al., "Motor theory of speech perception: A reply to Lane's critical review," (1970).
- <sup>13</sup>Bidelman, G.M., et al., "Tracing the emergence of categorical perception in the human auditory system," *NeuroImage*, 29, 201-212 (2013).
- <sup>14</sup>Bidelman, G.M. and B. Walker, "Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network," *NeuroImage*, 201, 116022 (2019).
- <sup>15</sup>Liebenthal, E., et al., "Specialization along the left superior temporal sulcus for auditory categorization," *Cerebral Cortex*, 20, 2958-2970 (2010).
- <sup>16</sup>Oldfield, R.C., "The assessment and analysis of handedness: The Edinburgh inventory," *Neuropsychologia*, 9, 97-113 (1971).
- <sup>17</sup>Bidelman, G.M., et al., "Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians," *European Journal of Neuroscience*, 40, 2662-2673 (2014).
- <sup>18</sup>Bakeman, R. and V. Quera, *Sequential analysis and observational methods for the behavioral sciences*. 2011: Cambridge University Press.
- <sup>19</sup>Oostenveld, R. and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clin. Neurophysiol.*, 112, 713-719 (2001).
- <sup>20</sup>Marie, C. and L.J. Trainor, "Development of simultaneous pitch encoding: infants show a high voice superiority effect," *Cerebral Cortex*, 23, 660-669 (2013).

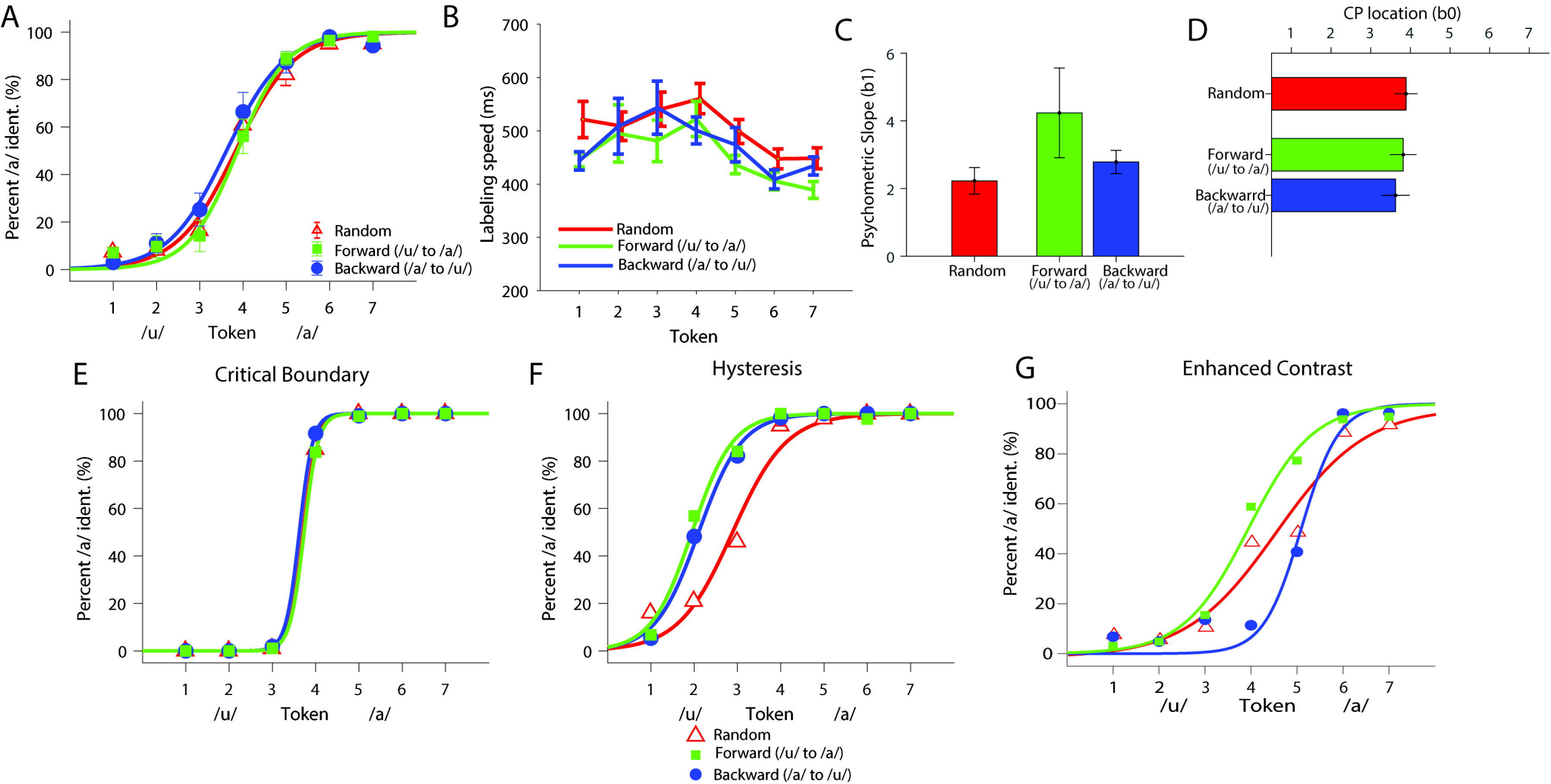


- <sup>21</sup>Maris, E. and R. Oostenveld, "Nonparametric statistical testing of EEG-and MEG-data," *Journal of neuroscience methods*, 164, 177-190 (2007).
- <sup>22</sup>Scherg, M., et al., "Taking the EEG back into the brain: the power of multiple discrete sources," *Frontiers in neurology*, 10, 855 (2019).
- <sup>23</sup>Pisoni, D.B. and J. Tash, "Reaction times to comparisons within and across phonetic categories," *Perception & Psychophysics*, 15, 285-290 (1974).
- <sup>24</sup>Reetzke, R., et al., "Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood," *Current Biology*, 28, 1419-1427 (2018).
- <sup>25</sup>Pisoni, D.B., "Auditory and phonetic memory codes in the discrimination of consonants and vowels," *Percept. Psychophys.*, 13, 253-260 (1973).
- <sup>26</sup>Pisoni, D.B., "Auditory short-term memory and vowel perception," *Mem. Cognit.*, 3, 7-18 (1975).
- <sup>27</sup>Altmann, C.F., et al., "Categorical speech perception during active discrimination of consonants and vowels," *Neuropsychologia*, 64C, 13-23 (2014).
- <sup>28</sup>Xu, Y., et al., "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," *J. Acoust. Soc. Am.*, 120, 1063-1074 (2006).
- <sup>29</sup>Bidelman, G.M., et al., "Effects of noise on the behavioral and neural categorization of speech," *Front. Neurosci.*, 14, 1-13 (2020).
- <sup>30</sup>Bidelman, G.M., et al., "Lexical influences on categorical speech perception are driven by a temporoparietal circuit," *J. Cogn. Neurosci.*, 33, 840–852 (2021).
- <sup>31</sup>Sayal, A., et al., "Identification of competing neural mechanisms underlying positive and negative perceptual hysteresis in the human visual system," *NeuroImage*, 221, 117153 (2020).
- <sup>32</sup>Moore, B.C.J., et al., "A model for the prediction of thresholds loudness and partial loudness," *Journal of the Audio Engineering Society*, 45, 224–240 (1997).
- <sup>33</sup>Iverson, P. and P.K. Kuhl, "Perceptual magnet and phoneme boundary effects in speech perception: do they arise from a common mechanism?," *Percept. Psychophys.*, 62, 874-86 (2000).
- <sup>34</sup>Healy, A.F. and B.H. Repp, "Context independence and phonetic mediation in categorical perception," *Journal of Experimental Psychology: Human Perception and Performance*, 8, 68 (1982).
- <sup>35</sup>Green, D.M. and J.A. Swets, *Signal detection theory and psychophysics*. 1966, New York,: Wiley. xi, 455 p.
- <sup>36</sup>Crowley, K.E. and I.M. Colrain, "A review of the evidence for P2 being an independent component process: age, sleep and modality," *Clinical neurophysiology*, 115, 732-744 (2004).
- <sup>37</sup>Key, A.P.F., et al., "Linking brainwaves to the brain: an ERP primer," *Developmental neuropsychology*, 27, 183-215 (2005).
- <sup>38</sup>Hickok, G., et al., "The role of Broca's area in speech perception: Evidence from aphasia revisited," *Brain and language*, 119, 214-220 (2011).
- <sup>39</sup>Bidelman, G.M. and B.S. Walker, "Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network," *Neuroimage*, 201, 1-10 (2019).
- <sup>40</sup>Fuhrmeister, P. and E.B. Myers, "Structural neural correlates of individual differences in categorical perception," *Brain Lang.*, 215, 104919 (2021).
- <sup>41</sup>Japee, S., et al., "A role of right middle frontal gyrus in reorienting of attention: a case study," *Frontiers in systems neuroscience*, 9, 23 (2015).

- 444 <sup>42</sup>Arsalidou, M., et al., "Brain areas associated with numbers and calculations in children: Meta-  
445 analyses of fMRI studies," *Developmental Cognitive Neuroscience*, 30, 239-250 (2018).  
446 <sup>43</sup>Zaleznik, E. and J. Park, "The neural basis of counting sequences," *Neuroimage*, 237, 118146  
447 (2021).  
448 <sup>44</sup>Kawahara, H., et al. "Tandem-STRAIGHT: A temporally stable power spectral representation  
449 for periodic signals and applications to interference-free spectrum, F0, and aperiodicity  
450 estimation". in *2008 IEEE International Conference on Acoustics, Speech and Signal*  
451 *Processing*. 2008: IEEE.  
452 <sup>45</sup>Nath, A.R. and M.S. Beauchamp, "A neural basis for interindividual differences in the McGurk  
453 effect, a multisensory speech illusion," *Neuroimage*, 59, 781-787 (2012).  
454 <sup>46</sup>Tools, P.S., *E-Prime Go*, P.S. Tools, Editor. 2020, Psychology Software Tools.

455

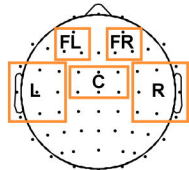




Random  
Forward  
Backward

Front left

Front right



Left

Center

Right

-300 nV

200 ms

