EFFECTS OF NONLINEAR DYNAMICS OF SPEECH CATEGORIZATION ON CORTICAL AND BRAINSTEM RESPONSES

Jared A. Carter

Follow this and additional works at: https://digitalcommons.memphis.edu/etd

Recommended Citation
MEMPHIS Graduate School

Final Committee Approval Form for Electronic Thesis or Dissertation Submission

This form serves as the official approval of an electronic thesis or dissertation by the student's faculty committee and must be submitted to the Graduate School with the defended and corrected copy for review. This form must not be included in the electronic document. This form must have original signatures of all committee members.

Degree Candidate's Name: Jared A. Carter

Degree (e.g., Ph.D): Ph.D __________________ Date of Defense: __________________

Major (e.g., Health & Sport Science): Communication Sciences and Disorders

Thesis/Dissertation Title: EFFECTS OF NONLINEAR DYNAMICS OF SPEECH CATEGORIZATION ON CORTICAL AND BRAINSTEM RESPONSES

We, hereby, certify that this document has been reviewed for form and content; and therefore, recommend that it be accepted for degree conferral.

Thesis or Dissertation Committee Names: 
(Please Print Name) 

Gavin M. Bidelman
Deborah Moncrieff
Eugene H. Buder
Stephanie Huette

Signatures:

Gavin M. Bidelman
Deborah Moncrieff
Eugene H. Buder
Stephanie Huette

I have reviewed the final version of the above-mentioned document and verify that it has been reviewed and accepted by the committee and is ready for electronic submission.

(Signature of Thesis/Dissertation Major Professor) 

26 May 2022
Graduate School
The University of Memphis
201 FedEx Institute of Technology Building, 365 Innovation Drive

Thesis/Dissertation Checklist

Listed below is a summary of the Graduate School on-line Thesis/Dissertation requirements. This checklist must be used in addition to what is specified by your committee/department. If your department required formatting changes different than the style manual or refereed journal, the Graduate School must be notified of these additional changes. This information must be included with your review document. Review Item 6 in the Thesis/Dissertation Preparation Guide for submission requirements of defended and corrected theses or dissertations.

This checklist must be submitted with your final document along with the Final Committee Approval Form for Electronic Thesis or Dissertation Submission. Both you and your major professor must check all applicable items followed for formatting purposes and each must sign. Please indicate the style guide used for formatting purposes: ☑ APA; ___ ASA; ___ Turabian; ___ Chicago Manual; ___ MLA; ___ LaTeX. If a journal was used, provide a sample article with your submission and indicate the name of the journal here: __________________________________________

MARGINS REQUIREMENTS:
☐ 1.00" left, top, right, and bottom margins for all pages. If you wish to eventually have bound copies, you may set your left margin for 1.50".

ORGANIZATION OF THESIS/DISSERTATION:
☐ 1. Title Page
☐ 2. Copyright page (Optional)
☐ 3. Dedication Page (Optional)
☒ 4. Acknowledgments (Optional)
☐ 5. Preface (Only when journal article formatting is used)
☐ 6. Abstract (350 words for dissertations, 150 words for theses)
☐ 7. Table of Contents
☐ 8. List of Tables (only used for 5 or more), with page numbers
☒ 9. List of Figures (only used for 5 or more), with page numbers
☐ 10. List of Plates (only if needed), with page numbers
☐ 11. List of Symbols and/or Abbreviations (only if needed), with page numbers
☑ 13. Bibliography/References/Works Cited
☐ 14. Appendix or Appendices
☒ 15. Permission letter(s) for any copyrighted materials used in text
☒ 16. IRB or IACUC Approval or waiver (if human or animal subjects were used – no signatures please)

FONT SIZE:
☐ 12 point is preferred; however, 10 point may be used, if approved by student's committee.

FONT TYPE:
☒ Times New Roman is preferred; however specialized fonts appropriate for specific typesetting needs (such as formulas and equations in Mathematical Sciences and other departments) may be used, if approved by student's committee.

SPACING:
☒ Double spacing will be the default for the entire document including before and after all centered headings and subheadings; this requirement may vary by department preference and/or by the style guide or refereed journal used for formatting.
PAGINATION:
- Every page must be assigned a number and page numbers should stand alone without any form of punctuation.
- All page numbers must be centered 1/2" from the bottom of the page (last line of text must be 1" from bottom).
- The title page is assumed to be numbered lowercase Roman numeral “i”; the actual number should NOT be displayed on this page. (Use a section break to solve this issue).
- All other preliminary pages (e.g., Copyright, Dedication, Acknowledgement, Abstract, Table of Contents, etc.) must be numbered in lowercase Roman numerals beginning with "ii".
- Pages in the body of text must be numbered using Arabic numerals beginning with "1".
- All page numbers in the Table of Contents and the List of Tables and List of Figures MUST correspond with actual page numbers in the text.

RUNNING HEADERS, FOOTNOTES AND ENDNOTES
- If footnotes are used, they must conform to margin requirements and must begin on the page they are cited.
- Footnotes must be two font sizes smaller than document text (10 pt).
- Single space footnote entries and double space between each.
- Footnotes are to be renumbered beginning with Arabic number "1" for each chapter or section.
- Running headers and endnotes are NOT allowed.

CENTERED HEADINGS AND SECTION HEADS
- Chapter headings MUST be centered; unless otherwise specified by student's committee. If they are not centered, please send justification with review copy.
- Section headings should be consistent throughout entire document.
- Preliminary page titles (i.e., Abstract, Table of Contents, etc.) and all chapter or section headings MUST be centered unless otherwise specified by student's committee.
- If a subheading falls at the end of a page without any accompanying text, move subheading to the next page.
- Do NOT use a numbering system for title and subheadings (e.g., 1.1, 1.1.1), unless required by style manual, refereed journal or by approval of student's committee. If they are numbered, please send justification with review copy.

TABLES/FIGURES:
- Format all tables/figures, including the caption, according to required style manual or refereed journal.
- All tables/figures must conform to required margin requirements.
- There must be a line before and after column headers and at the end of each table; however this may vary according to the style guide or journal used for formatting purposes.
- When a table is continued to another page(s), repeat table number and column headers; label table title as such - Table 1 (Continued)
- Tables/figures must be identified in the text by a number (e.g., Table 1, do not label as Table 1.1, Table 2.1, etc.), unless specified otherwise by student's committee.
- Table numbers and titles must be typed above the table.
- Figure numbers and captions must be typed below the figure.
- Table/figure captions must appear on the same page as the table/figure.
- If a table/figure is taken directly from another source, the entire source must be cited below the table/figure.
- Tables/figures may follow the page on which they are first referenced, or they may be included in a separate appendix. EXCEPTION: All tables/figures must be incorporated within the text for dissertations only.
- Tables/figures must be numbered consecutively throughout the text, unless specified otherwise by student's committee.

MISCELLANEOUS:
- Left align text (full justification of the right-hand margin is not allowed; margins must be jagged).
- If mathematical equations are used within the text, double space before AND after each equation.
- Always spell out a number if it begins a sentence or paragraph.
- Color figures, illustrations or charts are acceptable.
I have checked the manuscript for all of the above items.

Student's Name (please print): JARED A CARTER

Signature: [Signature] Date: 6/17/22

E-mail Address: jcrter29@memphis.edu

* I have checked the manuscript for all of the above items and represent committee approval of the selected style guide.

Thesis/Dissertation Chair’s Name (please print): Gavin Bidelman, PhD

Signature: [Signature]

Date: 6/16/22

Revised: 08/04/20
EFFECTS OF NONLINEAR DYNAMICS OF SPEECH CATEGORIZATION ON CORTICAL AND BRAINSTEM RESPONSES

by

Jared A. Carter

A Dissertation

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Major: Communication Sciences and Disorders

The University of Memphis

May 2022
Dedication

To my grandfather, thank you for initially encouraging me to obtain a Ph.D. To the rest of my family, thank for the love and support you have sent my way. To my D&D group, thank you for keeping me sane through this whole process and for the adventures had along the way.
Acknowledgements

Work in this dissertation was supported by a grant from the National Institute on Deafness and Other Communication Disorders (R01DC016267) awarded to Gavin M. Bidelman.
Preface

Abstract

Speech communication requires grouping continuous acoustic cues into meaningful phonetic units through categorical perception (CP). CP can be measured both behaviorally and neurally through electroencephalography (EEG). The perception of speech sounds is not static but can be modulated by stimulus context, such as by presenting token steps of a categorical continuum in sequential acoustical order (e.g., moving from low first formant (F1) frequency to high F1 frequency). Serial presentation order causes perceptual shifts, particularly around the categorical boundary, so an identical stimulus can be perceived differently by the same listener. Such perceptual warping of the acoustic-phonetic space can be described as a perceptual nonlinearity. This dissertation aimed to explore the neural mechanisms of these perceptual phenomena, especially when and where in the brain they occur, through the lens of speech categorization.

In complementary studies using EEG and behavioral measures, this dissertation evaluated top-down influences on cortical (Study 1) and subcortical (Study 2) speech processing, manipulating the sequencing of speech stimuli to bias listeners’ categorical percepts. In both studies, vowel stimuli were presented in random, forward (low to high F1), and backward (high to low F1) orderings as listeners categorized tokens as /u/ or /a/. Behaviorally, in both studies, listeners’ categorical boundaries shifted for individual listeners, indicating the effectiveness of the tasks in warping perception. Analysis of cortical event-related potentials (ERPs) showed that late (~300 ms) brain activity localized to superior temporal gyrus and middle/inferior frontal gyri predicted the magnitude of listeners’ perceptual warping effect (i.e., hysteresis/enhanced contrast).
Analysis of subcortical frequency-following responses (FFRs) revealed changes in the fundamental frequency (F0) amplitude indicative of category-level coding, as well as a biasing of responses to ambiguous tokens towards the listeners’ phonetic label. Both studies indicate top-down processing shapes the neural encoding and categorization of speech throughout the auditory system, including in the brainstem. Taken together, the results of this dissertation reveal the context surrounding the stimulus continuously shapes neural responses to speech in cortex and brainstem in a predictive, behaviorally-relevant manner. Such top-down modulation at multiple levels of speech representation might aid the perceptual system by reducing ambiguity in speech signals.
# Table of Contents

List of Figures .................................................................................................................. x

**Chapter 1: General Introduction** ................................................................................... 1

  Categorical Perception .......................................................................................................... 1

  Neural correlates of categorization at cortex and brainstem .............................................. 2

  Top-down effects and nonlinear dynamics of categorical perception ............................... 3

  Current Work ...................................................................................................................... 7

**Chapter 2: Nonlinear dynamics in auditory cortical activity reveal the neural basis of**

  perceptual warping in speech categorization ...................................................................... 8

  Methods ............................................................................................................................. 10

    Pilot experiment: Characterizing perceptual warping for vowels vs. CVs .................. 10

    Participants ...................................................................................................................... 10

    Stimuli & task ................................................................................................................. 11

    Behavioral data analysis ............................................................................................... 12

    EEG recording procedures and analysis ....................................................................... 14

  Results ............................................................................................................................... 16

    Behavioral data ............................................................................................................. 16

    Electrophysiological data ............................................................................................. 21

  Discussion ......................................................................................................................... 24

    Perceptual nonlinearities in categorization are stronger for vowels than CVs ............ 24
Perceptual warping of categories is subject to stark individual differences .......... 25

Electrophysiological correlates of perceptual warping........................................ 26

Chapter 3: Perceptual warping exposes categorical representations for speech in human brainstem responses ................................................................. 29

Introduction ............................................................................................................. 29

Materials & Methods .............................................................................................. 35

Participants .............................................................................................................. 35

Stimuli & Task ......................................................................................................... 35

EEG Recording Procedures ...................................................................................... 38

Behavioral Data Analysis ......................................................................................... 39

Electrophysiological Data Analysis ......................................................................... 40

Statistical Analysis .................................................................................................. 41

Results ..................................................................................................................... 42

Behavioral Data ....................................................................................................... 42

Electrophysiological Data ......................................................................................... 46

Brain-Behavior Relationships ..................................................................................... 52

Discussion ................................................................................................................. 53

FFR responses obtained concurrent with active task ................................................ 53

Brainstem FFRs are modulated by attention ............................................................ 55
Brainstem FFRs carry category-level information (perceptual correlates) of speech

Chapter 4: General Discussion

References

Appendix
List of Figures


**Figure 2.** 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 3.** 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/→/ɑ/ vs. 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; Fig. 2). Individual differences reveal unique forms of perceptual nonlinearity across sub-classes of listeners (n=3 representative subjects; one for each response pattern). (E) Critical boundary listener, where the individual selects the same response regardless of presentation order (s4). (F) Hysteresis listener, where the prior percept continues beyond the expected perceptual boundary (midpoint) as measured in sequential presentation (cf. panel E) (s1). (G) Enhanced contrast listener, where the
category response flips earlier than expected during sequential presentation (s2). See also

**Table 1** for classifications of these listeners. Error bars = ±1 s.e.m. .......................... 17

**Figure 4.** Grand average ERPs (at Tk4 = critical boundary stimulus) for forward, backward, and random presentation order of the vowel continuum. Boxes = channel electrode clusters. Negative voltage plotted up. ................................................................. 21

**Figure 5.** 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 channel cluster. 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., 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. ................................................................. 23

**Figure 6.** (A) Schematic of the presentation orders and examples of changes to the psychometric function the presentation order could create. (B) Schematic of the stimulus clustering paradigm for recording FFRs during active behavioral tasks [modified from Bidelman (2015c)]. 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 /a/. Following the behavioral response, a 400 ms pause occurred and the next block was presented. The clustered ISI sequence is then repeated to achieve the appropriate token counts for the FFR (x1000 presentations per token per condition) and sufficient behavioral responses (x50 per token).

**Figure 7.** Group level behavioral categorization responses. (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) 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.

**Figure 8.** Individual level psychometric functions. Representative subjects who showed (A) hysteresis (B) critical boundary and (C) enhanced contrast perceptual response patterns.

**Figure 9.** FFR time domain waveforms for Tk1 contrasting stimulus presentation orders and attentional state (i.e., active vs. passive listening).

**Figure 10.** FFR time domain waveforms comparing tokens Tk1, Tk4, and Tk7 (random condition).

**Figure 11.** FFR spectra (Tk1) in the backward, random, and forward conditions vs. the passive condition. Insets show F0 and F1 analysis windows.

**Figure 12.** 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 neither contrastive by
direction nor attention. (C) F0 amplitudes (pooling across orders) showed a U-shape
pattern suggesting of categorical coding (Pisoni, 1973). (D) F1 amplitudes (pooling
orders) were significantly larger for /u/ vs /ɑ/ ends of the continuum, which may have
related with stimulus acoustics. Errorbars = ±1 s.e.m. ........................................ 49

Figure 13. Brainstem FFRs show no adaptation in response to rapid speech trains.
Shaded areas=±1 s.e.m........................................................... 50

Figure 14. 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................................................................................... 52

Figure 15. Possible circuits for perceptual warping in speech categorization. Cortically,
medial frontal gyrus interacts with auditory cortex to shape perception. Subcortically, one
of two things could happen: (1) the modulation fed to auditory cortex further influences
subcortical responses or (2) medial frontal gyrus directly influences both auditory cortex
and subcortex. ........................................................................................................... 62
Chapter 1

General Introduction

Categorical Perception

Categorical perception (CP) is a phenomenon where stimuli with similar characteristics are grouped together in broad categories to facilitate perceptual-cognitive processing. CP has been observed in a wide variety of circumstances including color (Harnad, 1987; Repp, 1984), animals (Dale et al., 2007), and face perception (Beale & Keil, 1995). In speech perception, CP provides a window into how acoustic cues are mapped into phonetic categories in order to support speech communication (Liberman et al., 1967). In CP paradigms, a wide variety of acoustic differences including voice onset time (VOT) (Sharma & Dorman, 1999), formant transitions (Pisoni, 1973), and frication (Ferrero et al., 1982) can be morphed to form categorical continua, with equidistant steps between stimuli, yet having an abrupt change in percept between categories. This nonlinear transformation into phonetic groups is a hallmark of CP.

Behavioral measures of categorical perception include the psychometric function, which can be quantified by both the slope of the function ($\beta_1$) and the categorical boundary ($\beta_0$). Psychometric function slopes show the strength of categoricity across the continuum from one exemplar to an ambiguous token to the other exemplar. A steeper slope indicates the change from one category to the next happens more rapidly than if the slope was shallower (Bidelman, 2015a; Strouse et al., 1998; Xu, Gandour, et al., 2006). The categorical boundary indicates the point at which the psychometric function passes the 50% identification mark and indicates the acoustical space where the category shifts from one percept to another (Altmann et al., 2014; Ganong III & Zatorre, 1980).
Additionally, one can measure how rapidly a listener responds to 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, yielding an inverted U shape when plotting RTs across tokens of a continua (Pisoni & Tash, 1974).

**Neural correlates of categorization at cortex and brainstem**

Beyond behavioral measures, event-related brain potentials (ERPs) have been used to determine the neural bases of CP. When listeners perceive two sounds as categorically different, the ERPs begin to differentiate starting around the P2 wave (~200 ms after stimulus presentation), but not when the stimulus is within category (Dehaene-Lambertz, 1997). This perceptual difference persists even when the signal is acoustically the same, but perceived as different phonetic labels (i.e., an ambiguous token) (Bidelman et al., 2013). However, even within this time range (i.e., 200+ ms), not every ERP relates directly with categorization; while the P2 and P300 waves can reflect categorization, mismatch negativity (MMN) measures do not measure categorization, but rather acoustic differences between sounds (Maiste et al., 1995). When responses are obtained in passive listening conditions, the P2 does not change in amplitude, even if changes arise under active listening conditions (Alho et al., 2016; Bidelman & Walker, 2017; Carter, 2018). Collectively, these findings demonstrate CP modulates neural responses beginning at ~200 ms but only under active task demands (i.e., categorization is not automatic).

While these and other studies have established neural correlates of CP in cortical structures, whether acoustic-phonetic conversion occurs earlier in the auditory system (and is perhaps intrinsic) remains controversial (Bidelman et al., 2013; Kuhl & Miller,
Speech-evoked brainstem responses due not show categorical coding but rather follow properties of the stimulus acoustic (continuous coding scheme) (Bidelman et al., 2013). However, the lack of CP effects in brainstem responses may be due to different methodologies. Cortical responses are typically collected under active listening conditions, while subcortical responses are typically obtained under passive listening—which lack attention necessary for category structure (Alho et al., 2016; Bidelman & Walker, 2017; Carter, 2018). Cortical (and behavioral) responses also occur at different timescales and occur with much larger (5-10 x) amplitude than subcortical responses (Bidelman, 2014, 2015c; Cebulla et al., 2014; Ferm et al., 2013). Smaller evoked signals require more trials in the average to achieve sufficient signal-to-noise ratio (SNR) and thus may be harder to detect CP effects than with cortical responses (Hall III, 1992). Another issue with subcortical vs. cortical processing is differences in adaptation, with subcortical responses experiencing little change due to rapid stimulus presentation (i.e., fast inter-stimulus interval, ISI) (Ballachanda et al., 1992), compared to cortical responses which require much slower presentation speeds (i.e., longer ISI) for elicitation (Davis et al., 1966; Gilley et al., 2005). Collectively, these studies demonstrate the need for more optimal tasks to investigate possible categorical representations at the brainstem level.

**Top-down effects and nonlinear dynamics of categorical perception**

When information is encoded from the external world this encoding is propagated up a neural pathway from external stimuli to the cortex (i.e., bottom-up processing); however, these stimuli sometimes are modified by brain regions involved in cognitive functions (i.e., top-down processing) (Carter & Bidelman, 2021; Ganong III & Zatorre,
1980; Mankel et al., 2020; Sohoglu et al., 2012). While audition involves the transduction of acoustic signals to neuro-perceptual representations in a bottom-up fashion, top-down influences (e.g., from language context, expectancy) can influence how the signals arriving in the brain are interpreted. In a previous study, we examined whether such higher order processes (i.e., linguistic interference) influence the early-stage, acoustic-phonetic conversion processing of CP (Carter & Bidelman, 2021). We measured source level, event related brain potentials (ERPs) from auditory cortex (AC) and inferior frontal gyrus (IFG) as listeners rapidly categorized speech sounds along a /da/ to /ga/ continuum presented in three listening conditions: quiet, and in the presence of forward (informational masker) and time-reversed (energetic masker) 2-talker babble noise. We found that responses localized to auditory cortex (AC) showed top-down influences during CP tasks as evidenced by reduced ERP amplitudes for tokens categorized during informational masking relative to energetic masking, even when the masker is the same spectrally as a masker without lexical influences (i.e., time-reversal of the speech maskers). Our findings suggest, at least at a cortical level, the process of categorization involves an interplay between bottom-up and top-down influences as the brain converts acoustic (continuous) to phonetic (discrete) speech representations.

Another method that could be used to determine top-down and bottom-up interplay is through testing serial dependence in perception. Perceptual judgement and changes in psychometric pitch judgment using Shepard tones (i.e., complex tones with sine waves separated by octaves) found that tones in which the perceived pitch was equally likely as being labeled rising or falling in random presentation orders would shift percept based on prior stimuli (Chambers & Pressnitzer, 2014), including in biasing
neural responses based on the percept (Chambers et al., 2017). Further, when modelling pitch discrimination tasks, models that included prior response history performed more poorly than those that included it (Arzounian et al., 2017). These effects indicate presentation history influences current perception on pitch (i.e., fundamental frequency [F0]) perception, which suggests sequential presentation based on other formants might similarly have perceptual effects on current perception. This could indicate changes across categorical continua—largely driven by changes to formants—could yield ordering effects.

CP is overwhelmingly studied with presenting continuum stimuli in a random order and having listeners identify each token as it is presented. However, if one presents the stimuli in a sequential fashion (e.g., presenting tokens in order from left to right of the continuum), the categorical boundary can shift, suggesting perception can be differentially warped due to previously presented signals (Diehl et al., 1978; Healy & Repp, 1982). Perceptual warping around the categorical boundary provides another complementary window into the top-down influences on speech perception through the lens of nonlinear dynamical systems. This warping occurs when the history of previous stimuli influences the perception at or near the midpoint of the continuum to be drawn to one percept or the other. In other words, the perceptual space can become altered so the same stimulus can be perceived as different percepts (Nguyen et al., 2009; Tuller et al., 1994). In terms of CP, two possible phonemes exist at ends of a categorical continuum, and act as “attractors” (or perceptual magnets) in the perceptual landscape. As the landscape shifts by changing the strength of each attractor by presenting in a sequential fashion, each of these attractors provides a point of stability; however, one attractor will
be the dominant (i.e., more stable) percept for the token (Abraham et al., 1990; Kelso, 2012). Which percept dominates at any one point of the gradient depends on both the amount of sensory evidence and perceptual biasing at each stimulus token in three possible ways: (1) the boundary could remain constant in a critical boundary response pattern; (2) the percept could linger longer than would be expected in a hysteresis response pattern; or (3) the percept could switch earlier than would be expected in an enhanced contrast response pattern (Fig 1; (Tuller et al., 1994). Unfortunately, the neurophysiological mechanisms and circuitry underlying nonlinear dynamics in speech perception remain poorly understood.

**Figure 1.** Example of response patterns from Tuller, B., Case, P., Ding, M., & Kelso, J. A. S. (1994). The Nonlinear Dynamics of Speech Categorization. Journal of Experimental Psychology: Human Perception and Performance, 20(1), 3-16. [https://doi.org/0095-1523/94/S3.00](https://doi.org/0095-1523/94/S3.00). Reprinted with permission from publisher.
Current Work

To this end, the current dissertation aims to characterize the neural basis of top-down influences in speech perception through the unique lens of CP and related nonlinearities in perceptual processing. Two complementary studies will address how perceptual nonlinearities manifest at different stages of the auditory brain including cortical (Study 1) and subcortical (Study 2) levels.
Chapter 2

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

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 (Liberman et al., 1967; Pisoni, 1973). 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 (Kuhl, 1986). Alternatively, there is ample evidence that top-down processing influences speech categorization as suggested by enhancements observed in highly proficient listeners (Bidelman & Mankel, 2018; Bidelman et al., 2021; Carter & Bidelman, 2021; Ganong III & Zatorre, 1980) and biasing effects, when individuals hear a different category depending on the surrounding speech context (Diehl et al., 1978).

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) (Nguyen et al., 2009; Tuller et al., 1994; Tuller et al., 2008). Both stop consonant and vowel continua can produce context-dependent
shifts in perception, though stronger perceptual warping occurs with more ambiguous speech sounds (Studdert-Kennedy et al., 1970).

Event-related brain potentials (ERPs) have been used to examine the neural underpinnings of speech categorization (Bidelman et al., 2013; Bidelman & Walker, 2019; Liebenthal et al., 2010). 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 (Bidelman et al., 2013). 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 (Tuller et al., 1994). 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 (Nguyen et al., 2009; Tuller et al., 1994) and top-down influences in speech CP (Bidelman et al., 2021; Carter & Bidelman, 2021; Ganong III & Zatorre, 1980), we hypothesized (1) the location of listeners’ perceptual boundary would shift according to the direction of stimulus presentation (i.e., random vs. forward vs.
backward) and (2) perceptual warping would be accompanied by late modulations in the ERPs.

**Methods**

**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 (/da/ to /ga/) used in previous studies (Carter & Bidelman, 2021; Kawahara et al., 2008; Nath & Beauchamp, 2012). 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 (E-Prime Go, 2020).

**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) (Oldfield, 1971). Each gave written informed consent in compliance with the University of Memphis IRB.
Stimuli & task

We used a 7-token (hereafter “Tk1-Tk7”) vowel continuum from /u/ to /a/ synthesized in MATLAB 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 (/a/). We selected vowels over consonant-vowel (CV) syllables because pilot data suggested vowels were more prone to nonlinear perceptual effects (see Figure 2). We delivered stimuli binaurally through insert earphones (ER-3A) at 76 dB_A SPL. Sounds were controlled by MATLAB coupled to a TDT RP2 signal processor (Tucker-Davis Technologies, Alachua, FL).

![Psychometric functions](image)

**Figure 2.** 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.
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 /α/, 430 Hz to 730 Hz (i.e., vowel lowering), and (3) backward /α/ to /u/, 730 to 430 Hz (i.e., vowel raising). Forward and backward directions on such a continuum were expected to produce perceptual warpings (i.e., hysteresis) (Tuller et al., 1994). 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).

**Behavioral data analysis**

**Psychometric function analysis**

Identification scores were fit with sigmoid $P = 1/[1 + e^{-β1(x-β0)}]$, where $P$ is the proportion of trials identified as a given vowel, $x$ is the step number along the continuum, and $β0$ and $β1$ are the location and slope of the logistic fit estimated using non-linear least-squares regression (Bidelman & Walker, 2019; Bidelman, Weiss, et al., 2014). Leftward/rightward shifts in $β0$ location for the sequential vs. random stimulus orderings would reveal changes in the perceptual boundary characteristic of perceptual nonlinearity (Tuller et al., 1994). These metrics were analyzed using a one-way mixed-model ANOVA (subjects = random factor) with a fixed effect of condition (3 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, ~7% across all tokens/conditions/listeners) (Bidelman et al., 2013; Bidelman & Walker, 2019). RTs were analyzed using a two-way, mixed model ANOVA (subjects = random) with fixed effects of condition (3 levels: random, forward, and backward) and token (7 levels).

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 Querier program (v 5.1.23; Mangold International https://www.mangold-international.com/en/products/software/gseq). This compared listeners’ category labels for each continuum token (e.g., instances where say Tk 3 presentations were 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 (Bakeman & Quera, 2011). 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 1 for details.

**EEG recording procedures and analysis**

**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) (Oostenveld & Praamstra, 2001). 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.

**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) (Bidelman & Walker, 2019; Marie & Trainor, 2013). We used cluster-based permutation statistics (Maris & Oostenveld, 2007) implemented in BESA® Statistics 2.1 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. Lastly, we repeated this analysis for tokens 3-5, representing stimuli surrounding the categorical boundary where warping was most expected.

**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 waves to determine the intracerebral sources that account for perceptual non-linearities in speech categorization (Scherg et al., 2019). 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. 5A). Correlations between changes in β0 and CLARA activations evaluated which source regions predicted listeners’ perceptual warping of speech categories.
Results

Behavioral data

Psychometric function data

Listeners perceived vowels categorically in all three presentation orderings as evidenced by their sigmoidal identification functions (Fig. 3A). 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. 3C). 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. 3D).
Figure 3. 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/→/ɑ/ vs. 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; Fig. 2). Individual differences reveal unique forms of perceptual nonlinearity across sub-classes of listeners (n=3 representative subjects; one for each response pattern). (E) Critical boundary listener, where the individual selects the same response regardless of presentation order (s4). (F) Hysteresis listener, where the prior percept continues beyond the expected perceptual boundary (midpoint) as measured in sequential presentation (cf. panel E) (s1). (G) Enhanced contrast listener, where the category response flips earlier than expected during sequential presentation (s2). See also Table 1 for classifications of these listeners. Error bars = ±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. 3B). 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 vs. endpoints (Tk4 vs. Tk1/7: $p < 0.0001$), consistent with previous studies demonstrating slower RTs for category-ambiguous speech sounds (Bidelman & Walker, 2019; Carter & Bidelman, 2021; Pisoni & Tash, 1974; Reetzke et al., 2018). Pooling orders, comparisons between the left/right sides of the continuum (Tk1,2,3 vs. 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. 3D), perceptual nonlinearities were subject to stark individual differences (Fig. 3E-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 1 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.

**Cross-classification analysis**

**Table 1** displays 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.
Table 1: Yule’s Q values for Tk3/5 (i.e., tokens flanking the expected β0) and response patterns by participant. More negative/positive Yule’s Q denotes hysteresis/enhanced contrast response patterns, respectively. *Yule’s Q of medium-to-large effect size |Q| ≥ 0.33. †Individuals shown in Fig. 3E-G

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

**Figure 4** shows scalp ERP channel clusters to token 4 (critical stimulus at the perceptual boundary) across presentation orders (see Supplemental Figure for raw ERP data). Cluster based permutation tests (Maris & Oostenveld, 2007) 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. 5A**, 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 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. 5B**). 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 \(n\) within each subgroup.

**Figure 4.** Grand average ERPs (at Tk4 = critical boundary stimulus) for forward, backward, and random presentation order of the vowel continuum. Boxes = 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. 5C-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 $\beta_0$ location [$r=0.63$, $p=0.011$]. STG activations
did not correlate with behavior (corrected $p$’s > 0.05).
Figure 5. 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 channel cluster. 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., 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.
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 supplemental material) 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 (see supplemental material). Vowels are generally perceived less categorically than CVs (Altmann et al., 2014; Pisoni, 1973, 1975; Studdert-Kennedy et al., 1970). 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 (Bidelman et al., 2020; Xu, Gandour, et al., 2006). Vowel percepts are thus more ambiguous categorically, and consequently more susceptible to contextual influences and individual differences (Bidelman et al., 2021). 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 (Sayal et al., 2020; Tuller et al., 1994).

**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) (Moore et al., 1997). Alternatively, /a/ might have been heard as being a more prototypical vowel (i.e., perceptual magnet) (Iverson & Kuhl, 2000), perhaps owing to its higher frequency of occurrence in the English language (Hayden, 1950; Mines et al., 1978). 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 vs. falling pitch (Luo et al., 2007; Schouten, 1985). 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 (Bidelman et al., 2021; Carter & Bidelman, 2021) and decreases (i.e., speeds up) for more prototypical tokens (Pisoni & Tash, 1974). 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. (Diehl et al., 1978; Healy & Repp, 1982) 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 1; 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 (Nguyen et al., 2009; Tuller et al., 1994; Tuller et al., 2008). 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) (Green & Swets, 1966).

**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 (Crowley & Colrain, 2004; Key et al., 2005). 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) (Hickok et al., 2011). 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) (Bidelman et al., 2021; Bidelman & Walker, 2019; Carter & Bidelman, 2021). Related, we find direction-related modulations in the perceptual warping of speech categories (to an otherwise identical sound) are predicted by left IFG engagement. IFG engagement in our tasks is consistent with notions that frontal brain regions help shape behavioral category-level predictions at the individual level (Fuhrmeister & Myers, 2021). 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 (Japee et al., 2015). Additionally, it is active when holding information in working memory, such as performing mental calculations (Arsalidou et al., 2018), and has been implicated in processing ordered numerical sequences and counting (Zaleznik & Park, 2021). 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; (Bidelman & Walker, 2019; Carter & Bidelman, 2021)), top-down modulations from frontal lobes dynamically shapes category percepts online during speech perception.
Chapter 3

Perceptual warping exposes categorical representations for speech in human brainstem responses

Introduction

To effectively utilize language, individuals must convert continuous stimuli in the external world to phonetic category units (Goldstone & 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 & 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 & Miller, 1975) or if higher-order language and memory regions modulate category speech percepts in a top-down manner (Bidelman et al., 2021; Carter & Bidelman, 2021; Ganong III & Zatorre, 1980; Kuhl, 1986; Kuhl & 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, Gandour, et al., 2006). The categorical boundary indicates the point at which the psychometric function crosses 50% identification, marking the stimulus location where the category shifts from one percept to another (Altmann et al., 2014; Ganong III & Zatorre, 1980). Additionally, one can measure how rapidly a listener labels each token via reaction time (RT). RTs demonstrate the speed of speed processing, which increases (i.e., slows down) during more ambiguous or degraded tokens and decreases (i.e., speeds up) during more prototypical tokens, yielding an inverted U shape when plotting RTs across the continua (Pisoni & Tash, 1974).

CP is usually assessed by randomizing the presentation of stimuli in the categorical continuum. When presenting stimuli in their acoustical 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 & Repp, 1982), through perseverating the prior perception (i.e., hysteresis) or changing perception to the other category earlier than anticipated (i.e., enhanced contrast) (Tuller et al., 1994). This suggests the brain’s ongoing sorting of incoming acoustics into categorical phonetic representations is actively modulated in a top-down manner. 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 & 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 & Schreiner, 1988; Smith et al., 1975; Sohmer et al., 1977), and not cochlear origin (Skoe & 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 & Anderson, 1980; Sohmer et al., 1977). While most previous work has shown a subcortical origin of the FFR, recent neuroimaging studies have revealed cortical contributions to the response at low (~100 Hz) speech frequencies when recorded via magnetoencephalography (MEG) (Coffey, Herholz, et al., 2016). However, EEG work has convincingly demonstrated that subcortical structures (i.e., midbrain and even auditory nerve) provide the largest contribution to the scalp-recorded $\text{FFR}_{\text{EEG}}$ for most of the frequency bandwidth of speech (Bidelman, 2018b;
Bidelman & 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 & Picton, 2008; Krishnan, 2002; Skoe & Kraus, 2010). Given its unique time-frequency signature within the EEG, FFRs have been used to characterize subcortical processing of speech (Bidelman & Powers, 2018; Bidelman & Momtaz, 2021; Bidelman et al., 2013; Galbraith et al., 1995; Johnson et al., 2005; Musacchia et al., 2008; Russo et al., 2004; Skoe & Kraus, 2010) and musical sounds (Bidelman, 2013; Bones et al., 2014; Mankel & Bidelman, 2018), as well as track changes in neural encoding across the lifespan (Anderson et al., 2012; Bidelman et al., 2019; Bidelman, Villafuerte, et al., 2014; 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 show strong evidence of category structure. This is despite concomitant category representations observed in the same listeners’ cortical evoked potentials (Bidelman et al., 2013). However, categorization in most neuroimaging studies is tested under active tasks, while subcortical studies use passive listening, which reduces category coding in neural responses (Alho et al., 2016; Bidelman & Walker, 2017). Nevertheless, some evidence exists that show the possibility of category representation 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, Gandour, et al., 2006); 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, Krishnan, et al., 2006). 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, *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 FFR 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 & 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 & 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 & Weisz, 2019; Price & Bidelman, 2021) while
others finding little to no effect of attention on the FFR (Aiken & Picton, 2008; Dunlop et al., 1965; Galbraith & Kane, 1993; Varghese et al., 2015). If attention does influence the brainstem FFR, then actively categorizing speech during behavioral tasks should yield measurable changes 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 determine 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 categorical boundary measured behaviorally and thus successfully warped (biased) listeners’ percept. If brainstem speech coding is sensitive to categorization, we hypothesized FFRs would
show larger F1 amplitudes in sequential vs. random presentation orders. We also anticipated relationship between neural and behavioral measures if the FFR is indeed modulated by listeners’ ongoing categorical percept.

Materials & Methods

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 Study 1 (Carter et al., 2022). All spoke American English, 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.

Stimuli & Task

We used a 7-token vowel continuum from /u/ to /ɑ/. Each 100 ms token had a fundamental frequency of 150 Hz to avoid cortical contributions to the FFR which are restricted to low (< 100-120 Hz) stimulus frequencies (Bidelman, 2018b; Brugge et al., 2009). 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 with electrical shielding to prevent stimulus electromagnetic artifact from contaminating neural responses (Campbell et al., 2012; Price & 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 requiring minimally 1000 trials for response detection (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. 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 token counts to detect the FFR (x1000 presentations per token per condition) and sufficient behavioral responses (x50 per token) (see Figure 6 for paradigm).
Figure 6. (A) Schematic of the presentation orders and examples of changes to the psychometric function the presentation order could create. (B) Schematic of the stimulus clustering paradigm for recording FFRs during active behavioral tasks [modified from Bidelman (2015c)]. 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 /a/. Following the behavioral response, a 400 ms pause occurred and the next block was presented. The clustered ISI sequence is then repeated to achieve the appropriate token counts for the FFR (x1000 presentations per token per condition) and sufficient behavioral responses (x50 per token).

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 /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) (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 rule out
electromagnetic contamination of the neurophysiological recordings. This used an identical setup to the passive block only with the earphone removed from listeners’ ear canal thereby recording only “neural noise” (Price & Bidelman, 2021).

**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 standard for recording brainstem FFRs (Billings et al., 2019; Coffey, Colagrosso, et al., 2016; Gockel et al., 2013; Skoe & Kraus, 2010). This montage is used to obtain responses from the vertically oriented dipoles in the brainstem (Bidelman, 2015b; Chandrasekaran & Kraus, 2010). Interelectrode impedances were kept ≤6 kΩ. EEGs were digitized at 10 kHz. Responses were epoched (−5 – 105 window), artifact rejected (set to retain 95% of responses in each run) and averaged to derive FFRs for each vowel stimulus. Responses were then band-passed filtered (130 – 2000 Hz) for visualization and quantification. 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).
Behavioral Data Analysis

Psychometric Function Analysis

Identification scores were fit with sigmoid \( P = \frac{1}{1 + e^{-\beta_1(x-\beta_0)}} \), 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 & Walker, 2019; Bidelman, Weiss, et al., 2014). 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 & Walker, 2019). We included RTs \( \leq 250 \) 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 a percept during the ongoing token train.

Classifying Response Patterns

To classify participants based on their different listening strategies (hysteresis/enhanced contrast/critical boundary), we calculated the standard deviation of the categorical boundary across all listeners in the random condition. We then considered whether each individual’s categorical boundary fell within a 2SD around the group mean for the forward/backward. Participants whose boundary fell within 2SD were categorized as critical boundary listener. If their categorical boundary occurred before this window (e.g., Tk3 for backward; Tk5 for forward), we categorized them as hysteresis listener. If instead their categorical boundary happened after this window (e.g., Tk5 for backward;
Tk3 for forward), we categorized them as an enhanced contrast listener. See Study #1 for further details on these different listener subtypes.

**Electrophysiological Data Analysis**

**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. 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)].

**Adaptation Within Stimulus Trains**

To determine if neural adaptation occurred given the repetitive stimulus trains, we compared the F0 amplitude of the first and the last token in each train (for each continuum token: Tk1-Tk7). We compared the amplitude of the fundamental frequency between the first and the last tokens to determine if the neural responsiveness decreased with the rapid presentation of the stimuli within the trains. If adaptation occurred, it would indicate the fast repetition of speech stimuli stressed the auditory system (Pérez-
Adaptation might inadvertently account for differential amplitude changes with stimulus presentation order and confound our interpretations of hysteresis and categorical representations in the FFR.

*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, Tk5) (cf. Yellamsetty & 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 & 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. 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 & Bidelman, 2019).

*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 a 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 & 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.

**Results**

**Behavioral Data**

Listeners perceived the vowels categorically in all presentation orderings as seen in Figure 7. 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 ($F_{2,312} = 18.72, p < 0.001$). Categorical decisions were slower for backward versus forward ($p < 0.001$) 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 ($F_{6,312} = 1.41, p = 0.21$). This suggests that under our clustered stimulus paradigm, listeners may have decided the category while the stimulus train was still ongoing.
Figure 7. Group level behavioral categorization responses. (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.
While the group level categorical boundary was largely stagnant, individual-level data showed stark differences as a function of presentation order (Fig. 8). In the backward condition, most listeners retained a critical boundary \((n = 11)\), some showed enhanced contrast \((n = 4)\), and one showed hysteresis \((n = 1)\). In the forward condition, even more listeners demonstrated a critical boundary \((n = 13)\) with some listeners having enhanced contrast \((n = 3)\).
Electrophysiological Data

Figure 9 and Figure 10 show a sampling of time-domain waveforms for the FFR by condition and token respectively. These waveforms were analyzed in the frequency domain to determine differences in F0 and F1 frequency and amplitude.
Figure 9. FFR time domain waveforms for Tk1 contrasting stimulus presentation orders and attentional state (i.e., active vs. passive listening).

Figure 10. FFR time domain waveforms comparing tokens Tk1, Tk4, and Tk7 (random condition).
Figure 11 shows FFR spectra in response to Tk1 across stimulus orderings (random, forward, backward) and attention conditions. Figure 12 shows F0 and F1 measures more clearly. We found that F0 amplitude differed as a function of condition \( (F_{3,423} = 7.91, p < 0.0001) \) and token \( (F_{6,423} = 5.86, p < 0.0001) \). Post-hoc testing revealed the passive F0 amplitudes were smaller than all three active listening conditions (backward, \( p = 0.0004 \); forward, \( p = 0.0042 \); random, \( p = 0.0001 \)). The token effect was attributed to smaller F0 amplitudes in response to /u/ tokens (Tks 1-3) compared to /a/ tokens (Tks 5-7) \( (p < 0.0001) \). Conversely, F1 amplitude did not differ as a function of condition \( (F_{3,423} = 0.31, p = 0.8189) \), 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 critical, they indicate subcortical speech representations are enhanced with active attention.

Figure 11. FFR spectra (Tk1) in the backward, random, and forward conditions vs. the passive condition. Insets show F0 and F1 analysis windows.
Figure 12. 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 neither contrastive by direction nor attention. (C) F0 amplitudes (pooling across orders) showed a U-shape pattern suggesting of categorical coding (Pisoni, 1973). (D) F1 amplitudes (pooling orders) were significantly larger for /u/ vs /ɑ/ ends of the continuum, which may have related with stimulus acoustics. Errorbars = ±1 s.e.m.

We found FFRs across the categorical continuum displayed a quadratic trend for both F0 ($F_{6,423} = 131.79, p < 0.0001$) and F1 ($F_{6,423} = 131.79, p < 0.0001$) 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. These results indicate that, despite identical F0s in the stimuli, the FFR showed categorical coding of F0 only in sequential presentation orders (which elicited perceptual warping). In contrast, the quadratic trends were highly significant ($p < 0.0001$) for the F1 amplitudes in all conditions. These results suggest that the FFR F1 also showed categorical coding regardless of attention or presentation order.

**Figure 13** shows the F0 of the first and last tokens in each train to determine if neural adaptation occurred given the rapid, clustered nature of our stimulus presentation. We found no difference in F0 amplitude between the first and last response in each train ($F_{1,647} = 0.20, p = 0.6551$). This confirms there was little to no adaptation of brainstem responses due to the rapid succession of auditory stimuli (Bidelman & Powers, 2018) and thus rules the confound that serial order effects in the FFR data were due to mere neuronal fatigue.

![Figure 13. Brainstem FFRs show no adaptation in response to rapid speech trains. Shaded areas=±1 s.e.m.](image.png)
Figure 14 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).
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 ($R^2 = 0.63$). 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 ($R^2 =$

**Figure 14.** 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 /a/). Errorbars = ±1 s.e.m.
0.71). Evaluating individual terms revealed a significant predictor in FFR F0 frequency on listener’s psychometric slopes ($t_{11} = 2.87, p = 0.015$). These results suggest the degree to which subcortical responses code different speech features predict properties of listeners’ vowel categorization.

**Discussion**

The current study 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 this novel approach, we show that attention modulates the encoding of speech as early as the auditory midbrain and moreover, that brainstem FFRs encode speech categorically.

**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 & Picton, 2008; Bidelman et al., 2013; Skoe & 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 spurious. 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 & 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. Consequently, our data provide new and important evidence that speech-evoked brainstem responses, likely 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 that 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 the sound associates with. Additionally, the 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 & Tash, 1974). We did not observe this in the current study. This may relate to listeners deciding 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).

**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. 12). Attention effects in the FFR thus seems localized to low-frequency components of the speech signal (Holmes et al., 2018). The effect of attention on any property brainstem responses has been equivocal in previous work, with some studies supporting (Galbraith et al., 1998; Hartmann & Weisz, 2019; Price & Bidelman, 2021) and others refuting (Aiken & Picton, 2008; Dunlop et al., 1965; Galbraith & Kane, 1993; Varghese et al., 2015) attentional effects on FFRs.

Attentional modulation of the FFR 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 (Price & 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 identify from cortex further down the system, rending 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 (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 (see Carter et al., 2022 for similar evidence between frontal lobe and auditory cortex). 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 & Kane, 1993; Varghese et al., 2015) or oddball paradigms (Hartmann & Weisz, 2019; Price & Bidelman, 2021), which may allow listeners to periodically disengage from the task. 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 & Bidelman, 2021).

**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 *de novo* (i.e., 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 (Study 1 of this dissertation), 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 shown 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 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 where driven by the normal physiological byproducts of rapid auditory processing (i.e., refractory of neuronal firing). 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 & Llano, 2021; Lesicko & Geffen, 2022).

We also note that our speech stimuli were designed with F0s well above cortical phase locking limits (Aiken & Picton, 2008; Krishnan, 2002; Skoe & Kraus, 2010), which rules out the possibility that our FFR results are conflated by cortical contributions (Coffey, Herholz, et al., 2016). This, in addition to the lack of neural adaption (characteristic of more peripheral auditory nuclei), strongly supports a brainstem locus of our findings.

The F0 amplitudes in neural responses to /ɑ/ stimuli were larger than in responses to /u/ stimuli, unexpectedly revealing differences due to perception. These differences could be due to an overall offset of the root mean square amplitude for the response being demonstrated in the F0. Alternatively, changes in F1 have previously been revealed to increase F0 amplitude when F1 contains linguistically-relevant cues (Krishnan et al., 2011). We can rule out explanations due to loudness differences, as all tokens were matched in sound level and perceptual loudness (94.1±1.0 phon) (Moore et al., 1997). This suggests that acoustic differences do not underlie the differences we find in FFR F0 amplitude, but instead indicate that perceptual differences between sides of the continuum drive this neural differentiation. This difference in FFR F0 also cannot be attributable to acoustical F0 differences. In fact, an acoustic analysis showed the reverse pattern as in the FFR; stimulus F0 of the /u/ tokens was 250% higher than amplitudes of /ɑ/ tokens. Meanwhile, the FFR F0 amplitude to /u/ tokens were smaller, i.e., 47.9% the amplitude as /ɑ/ tokens. The dissociation of acoustics and neural responses indicates a neural, rather than acoustic, origin for the differences we observed. This view that
perception drives FFR differences is further bolstered by the similarities between neural responses to tokens within the same phonetic category.

Two alternative theories suggest how ambiguous phonemes are categorized and might account 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, partially consistent with NRV. It could be that the vowel biasing is more strongly related to NLM in our study as we were using 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 listeners’ categorical boundary and the
slope of their psychometric function. We and others have shown that perception begins to
differentiate phonemes categorically early in the cortical hierarchy and no later than
primary auditory cortex (Bidelman & Lee, 2015; Bidelman & Walker, 2019; Carter &
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). Collectively, the fact that FFRs are strongly modulated by
attention and show category-specificity (Fig. 12) 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.
Chapter 4

General Discussion

Using cortical and subcortical brain responses as well as behavioral responses, these studies further the work observing the top-down effects on the neural encoding and categorization of speech during nonlinear speech processing. Further this work increases our understanding of the effects of context on speech processing. The results from the first study indicate: (1) the warping of speech tokens was stronger for vowels than for CV syllables; (2) this warping effect was subject to strong individual differences; (3) the behavioral warping corresponded with late (~300 ms) neural effects over left hemisphere, with underlying sources in a frontotemporal circuit. These results indicate stimulus presentation order influences phonetic encoding, emerging from a top-down modulation of early auditory encoding by frontal brain regions. The results from the second study established a novel paradigm to obtain FFR and behavioral responses to speech concurrently, which showed: (1) attentional effects on audition as early as the auditory midbrain; and (2) categorical encoding of phonemes in brainstem FFRs. These results indicate the vital role auditory midbrain plays in the active perception and categorization of speech. Collectively, these studies reveal a brain network actively predicting speech based on prior input throughout the auditory system, with individuals engaging in different compensatory strategies to accurately perceive speech.

Through these studies, we have found a top-down neurophysiologic pathway by which speech can be categorized throughout the auditory system (see Fig. 15). Frontal regions associated with memory (i.e., medial frontal gyrus [MFG]) hold memory tracings of prior sounds and begin to predict what future sounds might be. These predictions are
sent down fronto-temporal and cortico-fugal pathways to shape perception of current sounds when the sound is ambiguous. Evidence from **Study #1** suggests the warping effect initiates in MFG, which modulates sensory encoding in auditory cortex (AC). **Study #2** demonstrates this effect influences sensory encoding in the early stages of auditory processing/pre-processing in the brainstem. To attempt disambiguating the relationships between MFG, AC, and brainstem, we ran correlations and linear models; however, none of the regions were significantly associated with any other; thus, we cannot parse out whether brainstem responses are modulated by MFG directly or via AC. Given that we previously found AC categorized sounds (Carter & Bidelman, 2021), we speculate the neural pathway underlying the warping of speech tokens would pass through AC; however further research should specifically investigate if the cortico-fugal pathways originate from auditory cortex or from frontal regions directly to subcortical FFR generators.

![Diagram](image)

**Figure 15.** Possible circuits for perceptual warping in speech categorization. Cortically, medial frontal gyrus interacts with auditory cortex to shape perception. Subcortically, one of two things could happen: **(1)** the modulation fed to auditory cortex further influences subcortical responses or **(2)** medial frontal gyrus directly influences both auditory cortex and subcortex.
This network seems involved with both tasks, despite the differences in how memory is interacting with phoneme perception. The first task involved a singular presentation per behavioral response; thus, the effect of memory came solely from the preceding stimuli. The second task retained that memory effect; however, further included echoic memory tracings through the rapid repetition of the phoneme for each behavioral trial. This would strengthen the memory tracing to solidify perception one way or the other before the listener responded (Recasens et al., 2015). Thus, individuals would have an initial perception, which the previous token modulated the initial percept, and that initial percept likely was strengthened over the course of the train, or else was quickly modulated by the other sounds in the train, long before the person behaviorally responds.

The differences between these two tasks led to differences in the results of both studies. However, despite these differences, both tasks yielded warping in perception. Study #1 had much longer RTs than Study #2. The median RTs for Study #1 were in the range of 400-600 ms, while Study #2 were in the range of 350-400 ms. These differences were likely due to the task design, with the first study requiring the response immediately after the listener categorized the stimulus, indicating identification speed. In contrast, the second study, the listener was primed for two seconds before the decision was made, so the RTs were less measuring identification speeds, and more detection speeds related to the end of the train. As there were no discernable differences in RTs across tokens in Study #2, using RTs to determine response patterns yielded no insights into listener’s response patterns using this metric.
In contrast to the differences in RTs, the psychometric function results were remarkably similar between the two studies. The slopes of the psychometric functions in sequential presentation orders were larger than in random for both studies, indicating that the strengthening of categories is not just a result of minor task design, but is driven by the prior stimuli influencing current perception (Tuller et al., 1994). Further, these studies both show there are no differences in the categorical boundaries for any condition at the group level. Collectively these findings suggest modifications in perception from higher-order brain regions cause warping of signals throughout the auditory system.

At the individual level, some data were similar between the studies, while other data were different. In common, both studies induce nonlinear warping of categorization between the directions at the individual level. Listeners are biased towards one end of the continuum vs. the other. However, in Study #1, cross-classification analysis revealed the nonlinearities, but in Study #2, cross-classification analysis failed to demonstrate nonlinearities. This failure was mainly due to the analysis technique requiring a minimum of N=5 observations in each row/column, and the limited number of behavioral data points (i.e., 50 behavioral responses per token) led to missing values in cross-classification analysis. Thus, for the second study, we used individual-level categorical boundaries in the random condition to get a confidence interval of ±2 standard deviations, and any boundaries beyond that range in sequential presentation were declared hysteresis/enhanced contrast (depending on the direction and the response). Reanalysis of Study #1 data using this method revealed in the forward direction n=13 critical boundary and n=2 enhanced contrast listeners. Meanwhile the backward direction had n=12 critical boundary listeners, n=1 enhanced contrast listeners, and n=2 hysteresis listeners. This
vast difference between the two methods of determining listener type suggest Study #2 might be underestimating warping in listeners. Thus, it is highly likely that hysteresis is the dominant mode of warping individuals use when processing sequentially presented speech tokens, based on the data from Study #1.

These studies further provided further evidence for the warping of speech perception to study the system. We demonstrated that memory areas (e.g., MFG) play a role in our perception of speech and that this effect modulates neural responses as early as auditory brainstem. Combined with our previous work with informational masking (Carter & Bidelman, 2021), we have added evidence that higher-order cognitive brain regions (e.g., language and memory regions) influence encoding at low-level sensory brain regions (e.g., auditory cortex and subcortex). Collectively, these findings indicate speech processing occurs in multiple levels of the brain and modulates perception of ambiguous tokens to streamline perception.

Future studies could seek to further our understanding of nonlinear dynamics in speech perception through combining our work in these studies with our prior speech-in-noise behavioral study (Bidelman & Carter, 2021) to determine the underlying neural mechanisms for perceptual warping improving phoneme recognition in noise. Further work could investigate the neural mechanisms underlying learning non-native phonemes (Tuller et al., 2008), to further work into the experience-driven changes at both cortical and subcortical levels. Additionally, while these studies have elucidated top-down influences on subcortical processing, it remains unclear if fronto-collicular or temporocollicular fibers drive this shaping in brainstem responses and deserves investigation.


https://doi.org/10.1097/AUD.0b013e31823fb585


negative perceptual hysteresis in the human visual system. *NeuroImage, 221*, 117153.


Appendix A

APA License for Copyrighted Materials

AMERICAN PSYCHOLOGICAL ASSOCIATION LICENSE
TERMS AND CONDITIONS

May 10, 2022

This Agreement between Mr. Jared Carter ("You") and American Psychological Association ("American Psychological Association") consists of your license details and the terms and conditions provided by American Psychological Association and Copyright Clearance Center.

License Number 5305430686828

License date May 10, 2022

Licensed Content Publisher American Psychological Association

Licensed Content Publication Journal of Experimental Psychology: Human Perception and Performance

Licensed Content Title The nonlinear dynamics of speech categorization.

Licensed copyright line Copyright © 1994, American Psychological Association

Licensed Content Author Tuller, Betty; Case, Pamela; Ding, Mingzhou; Kelso, J. A. Scott

Requestor type Author of requested content
<table>
<thead>
<tr>
<th>Format</th>
<th>Electronic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portion</td>
<td>Licensed Content Date</td>
</tr>
<tr>
<td></td>
<td>Licensed Content Volume</td>
</tr>
<tr>
<td></td>
<td>Licensed Content Issue</td>
</tr>
<tr>
<td></td>
<td>Type of Use</td>
</tr>
<tr>
<td></td>
<td>chart/graph/table/figure</td>
</tr>
<tr>
<td></td>
<td>Number of charts/graphs/tables/figures</td>
</tr>
<tr>
<td></td>
<td>Rights for</td>
</tr>
<tr>
<td>Title</td>
<td>Duration of use</td>
</tr>
<tr>
<td>Creation of copies for the disabled</td>
<td>no</td>
</tr>
</tbody>
</table>
Total 0.00 USD

With minor editing privileges no

In the following language(s) Original language of publication

With incidental promotional use no

The lifetime unit quantity of new product 0 to 499

Ph.D. Candidate

Institution name The University of Memphis

Expected presentation date May 2022

Portions Figure 2

Mr. Jared Carter
901 Telluride Drive

Requestor Location

ARLINGTON, TX 76001
United States
Attn: Mr. Jared Carter

Terms and Conditions

78
RightsLink Plus Terms and Conditions
American Psychological Association

The publisher for this copyrighted material is the American Psychological Association ("APA"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by the Copyright Clearance Center, Inc. ("CCC") at the time that you opened your CCC account.

APA hereby grants you a non-exclusive license to use this material, subject to the conditions below.
1. Limited License.
   Licenses are for one-time use only with a maximum distribution equal to the number
   that you identified in the licensing process; any form of republication must be
   completed with 12 months from the date hereof (although copies prepared before then
   may be distributed thereafter); and any electronic posting is limited to a period of 12
   months.

2. Credit Lines and Translation Disclaimer.
   The reproduced material must include one of the following credit lines:

   For English language content: Copyright © [year] by [Copyright Owner]. Reproduced
   [or Adapted] with permission. [Full bibliographic citation.]

   For translated language content: This material originally appeared in English as [full
   bibliographic citation]. Copyright © [year] by [Copyright Owner]. Translated and
   reproduced [or adapted] with permission. [Copyright Owner] is not responsible for the
   quality or accuracy of this translation. This translation cannot be reproduced or
   distributed further without prior written permission.

3. Author Permission.
   You must obtain the author's (or, in case of multiple authorship, one author's)
   permission. A good faith effort must be made to locate the author and obtain the
   author's permission. Author permission is not required for the reuse of small excepts
   or 1 to 3 figures or 1 to 3 tables; or for course pack or electronic reserve use.

   Licenses may be exercised anywhere in the world.

   If the APA content will be reproduced in a book, then permission is limited to the life
   of the current edition.

6. Online Use Limitations.
   APA full-text articles or book chapters may only be posted on a secure and restricted
   web site. The credit line specified in Section 2 shall appear on the first screen on
which the APA content appears and shall be amended to include the following additional wording: No further reproduction or distribution is permitted without written permission from the American Psychological Association.

7. Reservation of Rights.
APA reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC’s Billing and Payment terms and conditions.

8. License Contingent on Payment.
While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by APA or by CCC) as provided in CCC’s Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC’s Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and APA reserves the right to take any and all action to protect its copyright in the materials. "$0.00" fee means that APA waived the fee.

9. License Revocation.
APA or CCC may, within 60 days from the date of license, deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will APA or CCC be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to CCC for denied permissions.

10. Warranties.
APA warrants that it has, to the best of its knowledge, the rights to license reuse of this content. However, you must ensure that the material you are requesting is original to APA and does not carry the copyright of another entity (as credited in the published version). If any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that original source to reuse the material.

11. Indemnity.
You hereby indemnify and agree to hold harmless APA and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. No Transfer of License.
This license is personal to you, but may be assigned or transferred by you to your employer or a business associate if you give prompt written notice of the assignment or transfer to the APA. No such assignment or transfer shall relieve you of the obligation to pay the designated license fee on a timely basis (although payment by the identified assignee can fulfill your obligation).
13. No Amendment except in Writing.
   This license may not be amended except in a writing signed by both parties (or, in the case of APA, by CCC on APA's behalf).

14. Objection to Contrary Terms.
   APA hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and APA (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

15. Jurisdiction.
   This license transaction shall be governed by and construed in accordance with the laws of the District of Columbia. You hereby agree to submit to the jurisdiction of the federal courts located in the District of Columbia for purposes of resolving any disputes that may arise in connection with this licensing transaction.

Other Conditions:

v1.2
Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.
Appendix B
IRB Approval: Protocol #2370

Institutional Review Board
Division of Research and Innovation
Office of Research Compliance
University of Memphis
315 Admin Bldg
Memphis, TN 38152-3370

PI: Gavin Bidelman
Co-Investigator:
Advisor and/or Co-PI:
Department: IIS-Institute For Intelligent Systems
Study Title: Neural correlates of complex auditory perception
IRB ID: 2370
Submission Type: Renewal
Level of Review: Expedited

IRB Meeting Date:
Decision: Approved
Approval Date: April 15, 2020

Research Notes:
Findings:

The IRB has reviewed the renewal request. The University of Memphis Institutional Review Board, FWA00006815, has reviewed your submission in accordance with all applicable statutes and regulations as well as ethical principles.

Approval of this project is given with the following obligations:

1. If this IRB approval has an expiration date, an approved renewal must be in effect to continue the project prior to that date. If approval is not obtained, the human subjects consent form(s) and recruiting material(s) are no longer valid and any research activities involving human subjects must stop.
2. When the project is finished, a completion form must be completed and sent to the board.
3. No changes may be made in the approved protocol without prior board approval, whether the approved protocol was reviewed at the Exempt, Expedited or Full Board level.
4. Exempt approval are considered to have no expiration date and no further review is necessary unless the protocol needs modification.
5. Human subjects training is required every 2 years and is to be kept current at cifiprogram.org.

Thank you,
James P. Whelan, Ph.D.
Institutional Review Board Chair