Neuroscience 565 (2025) 182–191

Contents lists available at ScienceDirect

Neuroscience

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

Neural correlates of phonetic categorization under auditory (phoneme) and visual (grapheme) modalities

Gavin M. Bidelman [a,b,c,*], Ashleigh York [d,e], Claire Pearson [d]

[a] Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, IN, USA
[b] Program in Neuroscience, Indiana University, Bloomington, IN, USA
[c] Cognitive Science Program, Indiana University, Bloomington, IN, USA
[d] School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA
[e] Univeristy of Mississippi Medical Center, Jackson, MS, USA

ARTICLE INFO

Keywords:
Audiovisual processing
Categorical perception
electroencephalography (EEG)
Speech perception

ABSTRACT

This study assessed the neural mechanisms and relative saliency of categorization for speech sounds and comparable graphemes (i.e., visual letters) of the same phonetic label. Given that linguistic experience shapes categorical processing, and letter-speech sound matching plays a crucial role during early reading acquisition, we hypothesized sound phoneme and visual grapheme tokens representing the same linguistic identity might recruit common neural substrates, despite originating from different sensory modalities. Behavioral and neuroelectric brain responses (ERPs) were acquired as participants categorized stimuli from sound (phoneme) and homologous letter (grapheme) continua each spanning a /da/−/ga/ gradient. Behaviorally, listeners were faster and showed stronger categorization of phoneme compared to graphemes. At the neural level, multidimensional scaling of the EEG revealed responses self-organized in a categorial fashion such that tokens clustered within their respective modality beginning ~150–250 ms after stimulus onset. Source-resolved ERPs further revealed modality-specific and overlapping brain regions supporting phonetic categorization. Left inferior frontal gyrus and auditory cortex showed stronger responses for sound category members compared to phonetically ambiguous tokens, whereas early visual cortices paralleled this categorical organization for graphemes. Auditory and visual categorization also recruited common visual association areas in extrastriate cortex but in opposite hemispheres (auditory = left; visual = right). Our findings reveal both auditory and visual sensory cortex supports categorical organization for phonetic labels within their respective modalities. However, a partial overlap in phoneme and grapheme processing among occipital brain areas implies the presence of an isomorphic, domain-general mapping for phonetic categories in dorsal visual system.

Introduction

The seemingly trivial task of comprehending speech requires the brain to categorize incoming stimulus features into discrete chunks. Like most sensory phenomena, speech poses the problem of invariance: segments of continuous stimulus features must be mapped into discrete categories (Goldstone and Hendrickson, 2010). In speech perception, categorical processing can be inferred from tasks where listeners hear sounds along a morphed phonetic continuum (e.g., /ba/-/pa/) and are asked to label those sounds with a binary response. Typically, listeners perceive the same phoneme until reaching the midpoint of the continuum where their labeling abruptly shifts—the category boundary (Harnad, 1987; Liberman et al., 1967; Pisoni, 1973; Pisoni and Luce, 1987). The categorical division of the speech signal enables interpretation of the continuous acoustic stream as a sequence of phonemes that comprise words and form the basis of subsequent high-order linguistic units. And while originally thought to be unique to speech, several studies have shown that even non-speech stimuli are perceived in a categorial manner including music (Burns and Campbell, 1994; Burns and Ward, 1978; Howard et al., 1992; Klein and Zatorre, 2011; Locke and Kellar, 1973; Mankel et al., 2022; Siegel and Siegel, 1977; Zatorre and Halpern, 1979), colors (Fonteneau and Davidoff, 2007; Franklin et al., 2008), faces (Beale and Keil, 1995), and lines (Ferraro and Foster, 1986; Foster, 1983; Foster and Ferraro, 1989). Thus, both auditory and

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

https://doi.org/10.1016/j.neuroscience.2024.11.079
Received 25 July 2024; Accepted 30 November 2024
Available online 2 December 2024

visual features are supported by a categorical coding scheme in the brain's perceptual-cognitive system.

Language itself exerts strong influences on audiovisual category representations. For example, relative to English speakers, Chinese speakers show sharper categorization of Mandarin tones (Bidelman and Lee, 2015), and nonmusicians show sharper categorization for native speech sounds relative to unfamiliar musical sounds (Bidelman and Walker, 2017). Linguistic experience also influences categorization of visual information. For example, cultures that distinguish color terms lexically (e.g., shades of blue) show more precise categorization of color spectra (e.g., blue-green) (Winawer et al., 2007). The perception of visual Chinese characters also varies according to ones' experience with different Chinese orthographic writing systems (Yang and Wang, 2018). Language is therefore tightly coupled with perception and influences the categorical processing of sound and visual information alike.

In this vein, studies have shown that letter-speech sound integration is crucial for reading acquisition (Preston et al., 2016). Interestingly, more experienced readers show sharper perception of phonetic boundaries than poor readers (Mody et al., 1997; Werker and Tees, 1987) and there is robust evidence that the relationship between reading ability and phonological awareness is reciprocal—the ability to isolate phonemes improves with exposure to the alphabet, and reading ability improves with training on phoneme segmentation (for review, see Bentin, 1992). The link between reading and spoken language experience is therefore a critical component of language development. Humans acquire spoken language from birth, master its basic structure by age 3, and do not acquire written language until years later (Miller and Gildea, 1987). Given that it is a less sophisticated skill, early stages of the reading process must transform the visual input to a form that is compatible with the acoustic speech perception system to promote efficient comprehension (Godfrey et al., 1981). Presumably, this involves converting individual graphemes to phoneme-like internal representations, and then mapping sequences of graphemes to encoded representations of syllables and words (Liberman, et al., 1967).

Although grapheme and phoneme representations are both critical to language processing, auditory and visual categories might be subject to different degrees of categorical organization in the brain. One study found categorical processing of uppercase letters from a V/X continuum (Yasuhara and Kuklinski, 1978), suggesting linguistic experience with stimulus labels results in letters becoming perceptual wholes. However, other work has shown the perception of letters from a G/Q continuum changes more gradiently as a function of letter ambiguity rather than letter category (Massaro and Hary, 1986). Similarly, another study found that the peak in discrimination performance for an n/h continuum did not correspond with the category boundary (McIntyre and Di Lollo, 1991). These conflicting findings suggest visual graphemes might be perceived more continuously than their sound phoneme counterparts.

While studies of letter categorization have provided conflicting results, other studies have shown categorical processing of visual stimuli that are highly relevant to letter decoding—namely lines. Foster (1983) employed visual stimuli from a curved line continuum in a discrimination task in which participants identified which of four lines was different. The stimuli were employed in an identification task in which participants labeled items as 'straight', 'just curved', or 'more than curved'. Peaks in performance from the discrimination task correlated with perceived category boundaries in the identification task (Foster, 1983). Another study of line perception employed items from a curved line continuum in a three alternative forced choice task and found that discrimination performance peaked at the midpoint of the continuum, consistent with category representation (Ferraro and Foster, 1986). Similarly, Foster and Ferraro (1989) examined categorization for visual stimuli from a horizontal/vertical line offset continuum. They found that peaks in performance from a line position discrimination task correlated with category boundaries identified in a labeling task ('no gap', 'gap', 'more than just a gap'). These findings suggest the visual features inherent to letter objects (i.e., curved, horizontal, and vertical lines) are indeed perceived in a categorical manner. Letters vary along these and other visual dimensions (e.g., obliqueness) across various fonts and handwriting styles. Although like speech sounds, visual letters represent the basic elements of language and meaning, no studies have directly compared their perceptual categorization. Contrasting letter and speech processing could elucidate whether the neural mechanisms underlying phonological linguistic categories are specific to the input sensory modality.

To this end, the aim of the present study was to directly compare speech (phoneme) and letter (grapheme) perception to determine whether the brain employs analogous neural processes across modalities to map continuous features of auditory and visual signals into their discrete linguistic categories. As far as we are aware, no studies have directly compared phoneme (sound) and grapheme (visual) category mapping to identical phonetic units in a cross-modal design. To measure the categorical processing of letters and speech sounds, we recorded multichannel event-related brain potentials (ERPs) as listeners actively categorized stimuli along a "da-ga" phoneme and homologous "da-ga" grapheme continuum. Comparing ERPs to letters and speech sounds allowed us to (i) assess where/when linguistic category representations in each modality emerge in the brain and (ii) distinguish shared vs. segregated neural mechanisms in processing audiovisual analogues that share identical category identity.

Materials & methods

Participants

N = 16 young adults (3 male, 13 female; age: $M = 24.5$, $SD = 12.9$ years) participated in the experiment.[1] All exhibited normal hearing sensitivity confirmed via audiometric screening (i.e., <25 dB HL, octave frequencies 250–8000 Hz) and had normal or corrected-to-normal vision. Each participant was strongly right-handed (74.8 ± 27.0 % laterality index; Oldfield, 1971), had obtained a collegiate level of education (18.8 ± 2.7 years formal schooling), and was a native speaker of American English. On average, the sample had 3.25 ± 3.3 years of music training. All were paid for their time and gave informed consent in compliance with a protocol approved by the Institutional Review Board at the University of Memphis.

Auditory (phoneme) and visual (grapheme) stimulus continuum

We used a 5-step, stop-consonant /da/ to /ga/ sound continuum (varying in place of articulation) to assess CP for speech (e.g., Bidelman et al., 2019) (Fig. 1**A**). Each sound token (Tk) was separated by equidistant steps acoustically yet was perceived categorically from /da/ to /ga/. Stimulus morphing was achieved by altering the F2 formant region in a stepwise fashion using STRAIGHT (Kawahara et al., 2008). The original audio material for the /da/ and /ga/ endpoint exemplars were recorded by Nath and Beauchamp (2012). We chose a consonant–vowel (CV) continuum because compared to other speech sounds (e.g., vowels), CVs are perceived more categorically (Altmann et al., 2014; Pisoni, 1973) and carry more salient articulatory gestures and visual cues for perception (Moradi et al., 2017). All tokens were normalized in duration (350 ms), amplitude (75 dB SPL), and bandwidth (50–4000 Hz). The auditory stimuli were delivered binaurally through shielded insert earphones (ER-2; Etymotic Research) controlled by a TDT RP2 signal processor (Tucker Davis Technologies).

The visual grapheme continuum was created by morphing (5 steps) between the alphanumeric images of "da" and "ga." Visual morphing

[1] EEG was not recorded from one participant due to technical error resulting in a final sample size of n=15 for the neural data (behavioral data were unaffected). The current sample was identical to the participants reported in Bidelman et al. (2021).

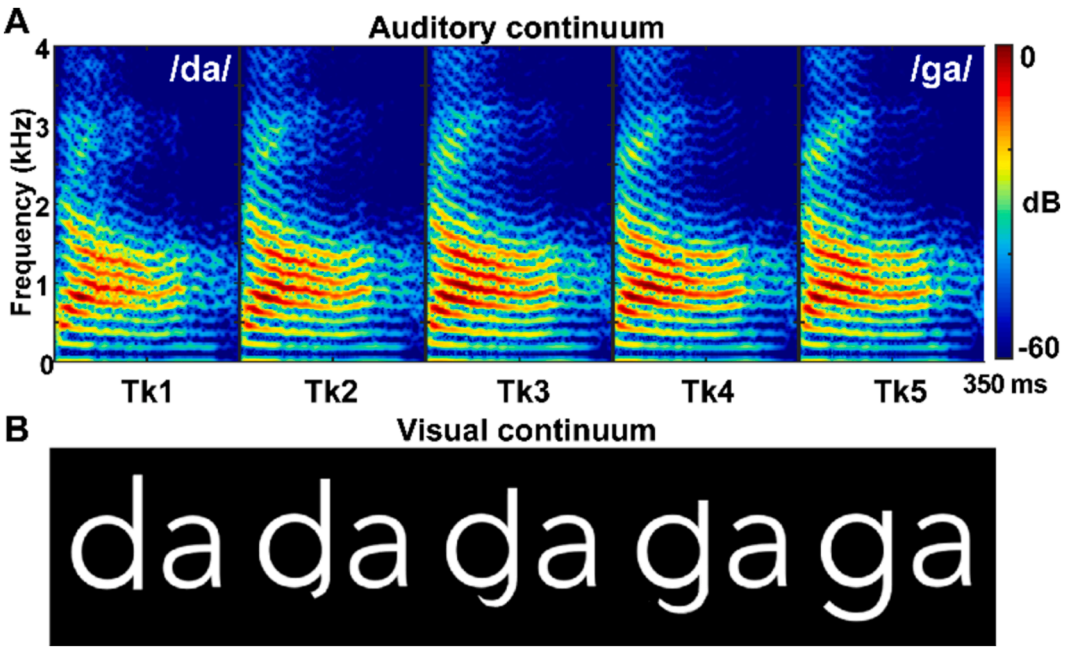

Fig. 1. Auditory and visual stimulus continua. (A) Phoneme sound continuum spanning 5 equidistant steps between "da" and "ga." Morphing was achieved by altering the F2 formant region in a stepwise fashion. (**B**) Visual grapheme continuum spanning 5 equidistant steps between "da" and "ga".

was achieved using custom scripts coded in MATLAB (e.g., Jonathan, 2011) (Fig. 1**B**).

During EEG recording, listeners heard or saw 150 trials of each individual token in auditory or visual blocks and labelled the stimulus with a binary response ("da" or "ga") as quickly and accurately as possible on the computer keyboard. Following, the interstimulus interval (ISI) was jittered randomly between 800 and 1000 ms (20 ms steps, uniform distribution) to avoid rhythmic entrainment of the EEG and anticipating subsequent stimuli. Block order for modality was randomized within and between participants. Visual stimuli (2.5" wide) were presented for 350 ms (matching the duration of the auditory stimuli) at the center of the computer screen (Samsung SyncMaster S24B350HL; nominal 75 Hz refresh rate) on a black background. The monitor was positioned at a distance of ~1 m resulting in a subtended visual angle of 4^0.

EEG recordings

EEGs were recorded from 64 sintered Ag/AgCl electrodes at standard 10–10 scalp locations (Oostenveld and Praamstra, 2001). Continuous data were digitized at 500 Hz (SynAmps RT amplifiers; Compumedics Neuroscan) using an online passband of DC-200 Hz. Electrodes placed on the outer canthi of the eyes and the superior and inferior orbit monitored ocular movements. Contact impedances were maintained <10 kΩ. During acquisition, electrodes were referenced to an additional sensor placed ~1 cm posterior to Cz. Data were re-referenced offline to the common average for analysis. Pre-processing was performed in BESA® Research (v7.1) (BESA, GmbH). Ocular artifacts (saccades and blinks) were corrected in the continuous EEG using principal component analysis (PCA) (Picton et al., 2000). Remaining trials exceeding ±150 μV were further discarded. Cleaned EEGs were then filtered (1–20 Hz), epoched (−200–800 ms), baselined to the pre-stimulus interval, and ensemble averaged resulting in 10 ERP waveforms per participant (5 tokens*2 modalities).

Behavioral data analysis

Identification scores were fit with a sigmoid function $P = 1/[1 + e^{-\beta 1(x\ -\ \beta 0)}]$, where P is the proportion of trials identified as a given phoneme, x is the step number along the stimulus continuum, and β_0 and β_1 the location and slope of the logistic fit estimated using nonlinear least-squares regression. Comparing parameters between speech contexts revealed possible differences in the "steepness" (i.e., rate of change) of participants' category labeling in the auditory and visual modality. Steeper functions represent stronger binary categorization. Behavioral labeling speeds (i.e., reaction times [RTs]) were computed as listeners' trimmed median response latency across trials for a given condition. RTs outside 250–2500 ms were deemed outliers (e.g., fast guesses, lapses of attention) and were excluded from the analysis (Bidelman et al., 2013; Bidelman and Walker, 2017).

EEG data analysis

ERP peak analysis. We measured the amplitude and latency of the auditory evoked potential (AEP) P2 deflection between 175–250 ms at the Cz electrode. We focus on the auditory P2 as we have previously shown this defection indexes auditory object and speech identification (Bidelman et al., 2020; Bidelman, et al., 2013; Bidelman et al., 2021; Bidelman and Walker, 2019) and tracks with perceptual learning during auditory categorization tasks (MacLean et al., 2024; Mankel, et al., 2022). Similarly, we measured the peak positivity from visual evoked potentials (VEPs) within the 375–450 ms time window at the PO8 electrode. These analysis windows were guided by visual inspection of the grand averaged data which showed peak activation in this timeframe (see Fig. 3).

ERP SNR. Comparisons between ERP classes might be spurious due to simple differences in signal quality (i.e., signal-to-noise ratio, SNR) between AEPs and VEPs. To rule out this possibility, we measured the SNR of each AEP and VEP per token. SNR was computed as the ratio of signal amplitude (see above) to the standard deviation within the post-stimulus epoch window (i.e., $\text{SNR} = \text{ERP}_{\text{amp}}/\sigma_{\text{epoch}}$), where σ_{epoch} is an estimate of noise overlapping with the evoked AEP/VEP (Bidelman et al., 2018; Hu et al., 2010). Critically, ERP SNR did not differ between modality [$F_{1,126} = 2.03$, $p = 0.16$] nor token [$F_{4,126} = 0.62$, $p = 0.63$], indicating that AEP and VEP responses were not inherently noisier than one another or between tokens.

Topographic ANOVA (TANOVA). To provide a more comprehensive analysis of where effects emerged over the entirety of time and space, we

used a topographic ANOVA (TANOVA) to identify the spatiotemporal points where the ERPs were sensitive to our stimulus manipulations (i.e., token and modality) (for details, see Bidelman and Yellamsetty, 2017; Koenig and Melie-Garcia, 2010; Murray et al., 2008). TANOVAs were implemented in the MATLAB package Ragu (Habermann et al., 2018; Koenig et al., 2011).

The TANOVA used a randomization procedure (N = 500 resamples) that tested the distribution of the ERP's topography in the measured data against a surrogate distribution, derived by exchanging all conditions and electrodes in the data. The percentage of shuffled cases where the effect size obtained after randomization was equal to or larger than the measured effect size obtained in the observed data provided an estimate of the probability of the null hypothesis. This analysis yielded running *p*-values across the epoch that identified the time samples at which the ERPs were significantly modulated by main (modality, token) and/or interaction effects (modality x token). To be considered a reliable effect (and prevent Type I error inflation), the procedure required a duration threshold whereby ≥46 ms of contiguous samples had to survive a $p < 0.05$ criterion to be considered a significant time window.

From the running TANOVA, we identified time segments where the ERPs showed modulations with both stimulus factors (i.e., modality x token interaction). Within these significant windows, differences in factor levels were visualized by computing ERP difference maps (averaged across the window's duration) between the scalp topographies for each condition. Because the number of case dimensions is large (e.g., 64 sensors x 2 modalities x 5 tokens x 15 subjects), the visualization was reduced to a two-dimensional space using a multidimensional scaling (MDS) approach (Koenig, et al., 2011). Similarities between the mean scalp topographies of the different conditions were assessed using the covariance between maps. The two-dimensional space that optimally represented the covariance matrix was represented in the first two eigenvectors. MDS visualization was then achieved by projecting the mean scalp map for a given condition/factor level onto the two eigenvectors, yielding a set of two-dimensional coordinates of each mean different scalp field map that was displayed as a scatterplot (see Fig. 5B). Points closer in the MDS neural space indicate a high degree of similarity between the scalp topographies whereas farther points reflect more dissimilar neural responses (Bidelman, et al., 2013).

Source imaging analysis. To estimate the underlying sources contributing to categorial processing, we used Standardized Low Resolution Electromagnetic Tomography (sLORETA) (Pascual-Marqui, 2002) to estimate the neuronal current density underlying the scalp ERPs. This distributed inverse method uses a standardized, unweighted minimum norm. sLORETA models the inverse solution as a large collection of elementary dipoles distributed over nodes on a mesh of the cortical volume. The algorithm estimates the total variance of the scalp data and applies a smoothness constraint to ensure current changes minimally between adjacent brain regions (Michel et al., 2004; Picton et al., 1999). sLORETA source images were computed in the 154–256 ms time window, where modality x token interaction effects were prominent in the TANOVA of the scalp data (see Fig. 5).

Results

Behavioral identification functions are shown for phoneme and grapheme continua in Fig. 2. Listeners showed stair-stepped labeling in both stimulus modalities confirming a sharp flip in their category percept from "da" to "ga" at the midpoint of each continuum (Fig. 2**A, B**). A 1-way ANOVA revealed identification slopes were steeper for auditory compared to visual stimuli [$F_{1,14} = 90.11$, $p < 0.0001$; $\eta_p^2 = 0.87$] indicating stronger categorical hearing of sounds than homologous visual tokens.

RTs were highly sensitive to both stimulus manipulations. Decision speeds showed strong effects of modality [$F_{1,126} = 424.04$, $p < 0.0001$; $\eta_p^2 = 0.2077$ and token [$F_{4,126} = 9.53$, $p < 0.0001$; $\eta_p^2 = 0.23$], but more critically, a modality x token interaction [$F_{4,126} = 7.74$, $p < 0.0001$; $\eta_p^2 = 0.20$] (Fig. 2**C**). The interaction was attributed to a slowing of RT speeds near the midpoint of the continuum where category membership becomes perceptually ambiguous (Bidelman and Carter, 2023; Bidelman and Walker, 2017; Pisoni and Tash, 1974). This inverted V-shape pattern in RTs was observed only for the V (contrast Tk3 vs. mean of others; $t_{126} = 6.84$, $p < 0.0001$) but not the A continuum ($t_{126} = 0.413$, $p = 0.68$). Taken together, the overall sharper categorical pattern and faster overall RTs for A vs. V tokens suggests stronger categorization for auditory phoneme compared to visual grapheme stimuli.

Grand average auditory- (AEPs) and visual-(VEP) evoked potentials elicited by phoneme and graphemes, respectively, are shown in Fig. 3. AEPs revealed a canonical P1-N1-P2 response with frontocentral topography that flipped in polarity at the mastoids—consistent with neural generators in the supratemporal plane (Picton, et al., 1999). Token-related modulations were observed in the time window of the P2 (~200 ms), consistent with the notion this wave reflects auditory object and speech categorization (Bidelman, et al., 2020; Bidelman, et al., 2013; Bidelman, et al., 2021; Bidelman and Walker, 2019; MacLean, et al., 2024; Mankel, et al., 2022). In contrast, VEPs were maximal at the posterior of the scalp, consistent with generators in the visual cortices surrounding the calcarine fissure (Ducati et al., 1988). VEPs showed stimulus-related modulations first at ~250 ms and peaking by 300–400 ms after grapheme onset.

A mixed-model ANOVA conducted on ERP peak amplitudes showed main effect of modality [$F_{1,126} = 4.78$, $p = 0.03$; $\eta_p^2 = 0.04$] (Fig. 4**A**). ERP latencies showed a main effect of stimulus modality that paralleled the behavioral RTs. VEPs were later than their AEP counterparts across the board [$F_{1,126} = 3154.09$, $p < 0.0001$; $\eta_p^2 = 0.96$]. However, this effect was expected given the difference in analysis window and peak deflection across AEP and VEP waveforms. More critically, ERP latencies showed a modality x token interaction [$F_{4,126} = 2.62$, $p = 0.038$; $\eta_p^2 = 0.08$] (Fig. 4**B**). The interaction suggests that the degree of categorical coding across tokens varied by stimulus modality. Supporting this notion, a repeated measures correlation (rmCorr) (Bakdash and Marusich, 2017) across tokens and modalities showed ERP latency was positively correlated with behavioral RTs ($r_{rm} = 0.82$, $p < 0.0001$). Faster neural latencies were associated with faster categorization speeds, which was more prominent in the auditory than visual modality.

TANOVAs (Koenig, et al., 2011; Murray, et al., 2008) conducted on the ERP topographies confirmed significant modulations in evoked activity with changes in both stimulus modality and token, as well as segments sensitive to both acoustic factors (i.e., modality x token interaction) (Fig. 5**A**). By itself, the token effect modulated activity across the middle portion of the response time course beginning at ~300 ms post stimulus onset. The main effect of modality was more pervasive, with significant modulations beginning as earlier as ~150 ms. Source imaging showed modality specific activations in this early time window (138–254 ms; see †) confirming that auditory phonemes activated auditory superior temporal cortex and visual graphemes activated occipital lobe (see Fig. 6), respectively. More critical was the token x modality effect. This interaction effect was circumscribed to a similar early (~154–256 ms) and additional late (~700 ms) time window after stimulus onset. Because this later (700 ms) window encroached on the behavioral RTs (see Fig. 2) it is confounded by post-perceptual processing and motor activity that appears late in the time-course of categorization tasks.[2] As such, we focused subsequent analysis on the early (154–256) interaction window that reflects sensory-perceptual encoding (rather than decision or motor planning) during speech categorization (Mahmud et al., 2021).

[2] Confounding motor activity was also confirmed empirically in sLORETA source imaging of the late window (~700 ms) (data not shown), further justifying its exclusion from the analysis.

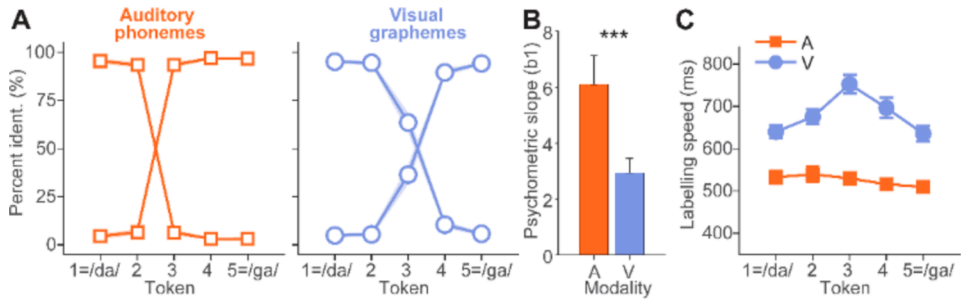

Fig. 2. Behavioral categorization of auditory phonemes and visual graphemes of a /da/-/ga/ continuum. (A) Identification functions. A and V token labeling shows a stair-stepped identification function consistent with categorical hearing; listeners perception abruptly flips at the midpoint of the continua (i.e., categorical boundary). (B) Slopes of the psychometric functions for auditory and visual categorization. Participants showed sharper (i.e., more categorical) perception of A vs. V tokens. (C) Reaction time (RT) speeds for token labelling. Responses are faster for A than V stimuli overall. Visual stimuli also evoked a slowing near the midpoint vs. endpoint tokens, consistent with category ambiguity near the midpoint of the continuum (Pisoni and Tash, 1974). Errorbars = ±1 s.e.m., ***$p < 0.0001$.

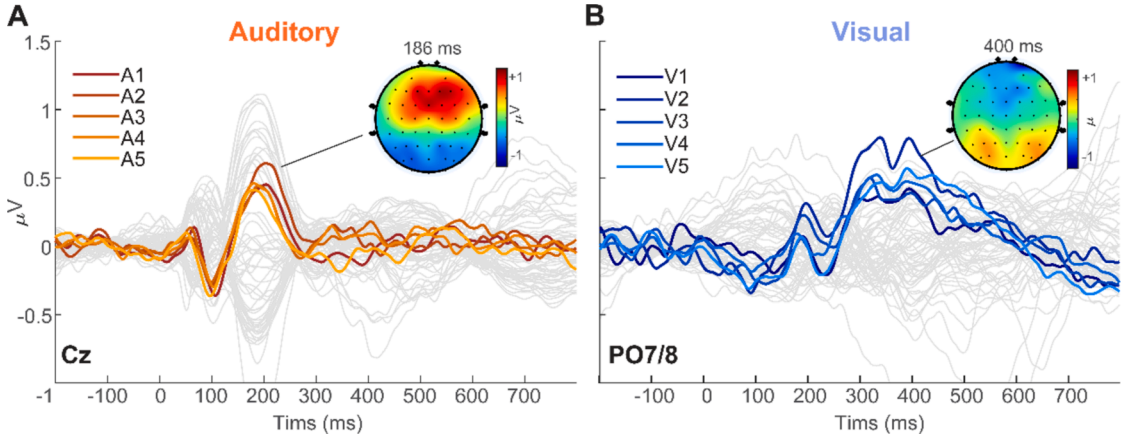

Fig. 3. Grand average cortical brain responses to phoneme and grapheme stimuli along a /da/-/ga/ continuum. (A) AEPs. Gray lines = 64 electrodes; colored lines = Cz electrode. Auditory responses reveal a canonical P1-N1-P2 response with frontocentral topography. Note the token-related modulations in the time window of the P2 (~200 ms) (**B**) VEPs. Gray lines = 64 electrodes; colored lines = mean of PO7/PO8 electrodes. VEPs showed stimulus-related modulations first at ~ 250 ms, peaking by 300–400 ms. Topographic maps are pooled across tokens for each modality.

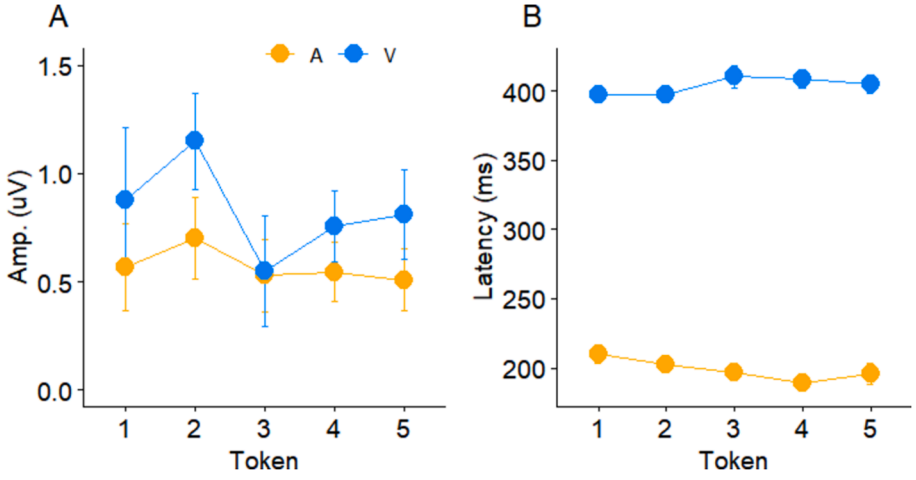

Fig. 4. ERP (A) amplitude and (B) latency as a function of stimulus modality and CV token. Measurements were taken at the Cz (AEP) and PO8 (VEP) electrodes. Errorbars = ±1 s.e.m.

MDS visualization of the AEP and VEP responses to phoneme and grapheme stimuli are shown in Fig. 5**B**. Responses clustered into two distinct "clouds" in the MDS based on stimulus modality (A vs. V: DIM #1). Similarly, individual tokens showed differentiation along the orthogonal DIM #2. Consistent with prior studies examining the neural differentiation of phonetic categories in AEPs (Bidelman and Lee, 2015; Bidelman, et al., 2013; Chang et al., 2010), within-category tokens (e.g., Tk1/2) tended to cluster in closer proximity to one another in the MDS space but far from their across-category counterparts (e.g., Tk 4/5). More critically, tokens grouped within their respective modality, as

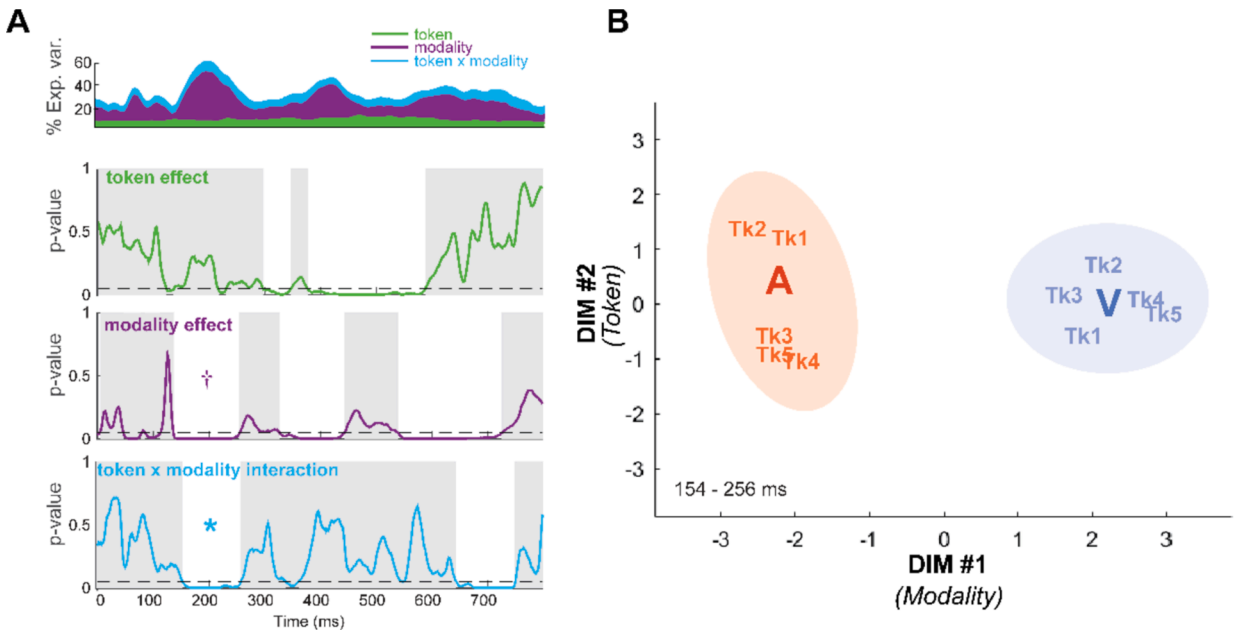

Fig. 5. Topographic ANOVA (TANOVA) revealing the time course at which the brain distinguishes phoneme and grapheme tokens. (**A**) *Top,* running time course of the %-explained variance of the TANOVA over all electrodes and time for the main and interaction effects. *Bottom 3 rows*, time course for the main effect of token, modality, and token x modality interaction. Each trace represents the running *p*-value for the effect, computed via permutation resampling (N = 500 shuffles). Dotted lines mark the $p = 0.05$ significance level. shaded regions = *n.s.* † = early modality main effect [138–254 ms] (see Fig. 6). * = significant token x modality interaction [154–256 ms] (see Fig. 7). (**B**) MDS visualization of the early sensory token x modality interaction (i.e., see* panel A). MDS visualization of the differences between neural responses to phoneme and grapheme stimuli. Note the clear separation of responses to A and V tokens along dimension #1 (*modality*) and clustering of within-category tokens (e.g., adjacent Tk1/Tk2 and far from Tk4/Tk5) within each modality. Category clustering of phonemes/graphemes appears more prominent in the auditory than visual modality.

indicated by the token x modality interaction observed in the TANOVA. However, category clustering of phonemes/graphemes appeared to be more prominent in the auditory than visual modality as indicated by a tighter convergence of within-category stimuli and farther separation across categories, respectively.

To resolve the neuronal sources underlying these modality-specific responses during categorization (i.e., interaction effect in Fig. 5A), we used sLORETA imaging (Pascual-Marqui, 2002) to visualize the current densities on the cortical surface. Statistical maps were generated contrasting the degree of categoricity in the neural AEP and VEP responses. To this end, we first computed difference waves between the two prototypical (i.e., mean Tk1/5) and phonetically ambiguous (i.e., Tk 3) tokens (separately for A and V responses). These difference waves index the degree of categorial coding in the ERPs (Bidelman and Walker, 2017; Bidelman and Walker, 2019; Liebenthal et al., 2010). We then calculated *t*-statistic maps contrasting this categorical coding index between modalities [i.e., $(A_{Tk1/5} - A_{Tk3}) - (V_{Tk1/5} - V_{Tk3})$]. Maps were computed in the 154–256 ms time window, where responses showed a maximal token x modality interaction (see *, Fig. 5A). The resulting statistical maps compared the degree of categorical coding in the auditory vs. visual modality across the entire brain volume (not raw sensory-specific activations as in Fig. 6).

Differential source activations for phoneme vs. grapheme category coding are shown in Fig. 7. Category coding for speech sound phonemes was stronger than visual graphemes in bilateral auditory cortex (AC), left inferior frontal gyrus (IFG), and middle frontal gyrus (MFG). Surprisingly, auditory categorical coding more strongly recruited portions of left occipital cortex including extrastriate visual association areas (BA 18/19). In contrast, visual categories recruited a network involving nodes in bilateral precentral gyrus (PCG), cuneus (CUN), and the right dorsal stream of the visual pathway including lateral occipital cortex and visual association areas (BA 18/19).

Discussion

A key aspect of language comprehension is the ability to categorize both acoustic and written inputs to form discrete linguistic units. Understanding how different sensory modalities perform this mapping between stimulus features and abstract linguistic space is important for understanding human speech comprehension. By measuring behavioral and neuroelectric brain responses (ERPs) during a sound (phoneme) and letter (grapheme) /da/ – /ga/ continuum, our data reveal (i) both modality-specific and overlapping brain regions support cross-modal categorization and (ii) stronger categorization for speech phonemes than their homologous orthographic counterparts. Collectively, our results imply that acoustic information enjoys a privileged role in the perceptual-cognitive operation of categorization.

Categorization is more salient for auditory than visual linguistic stimuli

Behaviorally, we found both phonemes and graphemes elicited the typical, stair-stepped identification functions characteristic of categorical hearing (Pisoni, 1973). Whether or not letters are perceived categorically has been somewhat equivocal in the literature (Massaro and Hary, 1986; Yasuhara and Kuklinski, 1978). Our data extend prior studies on visual objects (e.g., lines, colors, and faces; Beale and Keil, 1995; Ferraro and Foster, 1986; Fonteneau and Davidoff, 2007; Foster, 1983; Foster and Ferraro, 1989; Franklin, et al., 2008) by confirming a similar category mapping for orthographic CV letters. However, we show here that both auditory and visual CV homologues are perceived in a categorical manner but to differing degrees. Listeners were faster and showed stronger, more discrete labeling of phoneme compared to grapheme CV tokens. Moreover, labeling was slower overall for visual than auditory stimuli and graphemes showed an additional slowing near the continuum's midpoint which was not observed for phoneme tokens. Our identification and RT data suggest that visual graphemes were perceived less categorically (i.e., more continuously) and/or were more

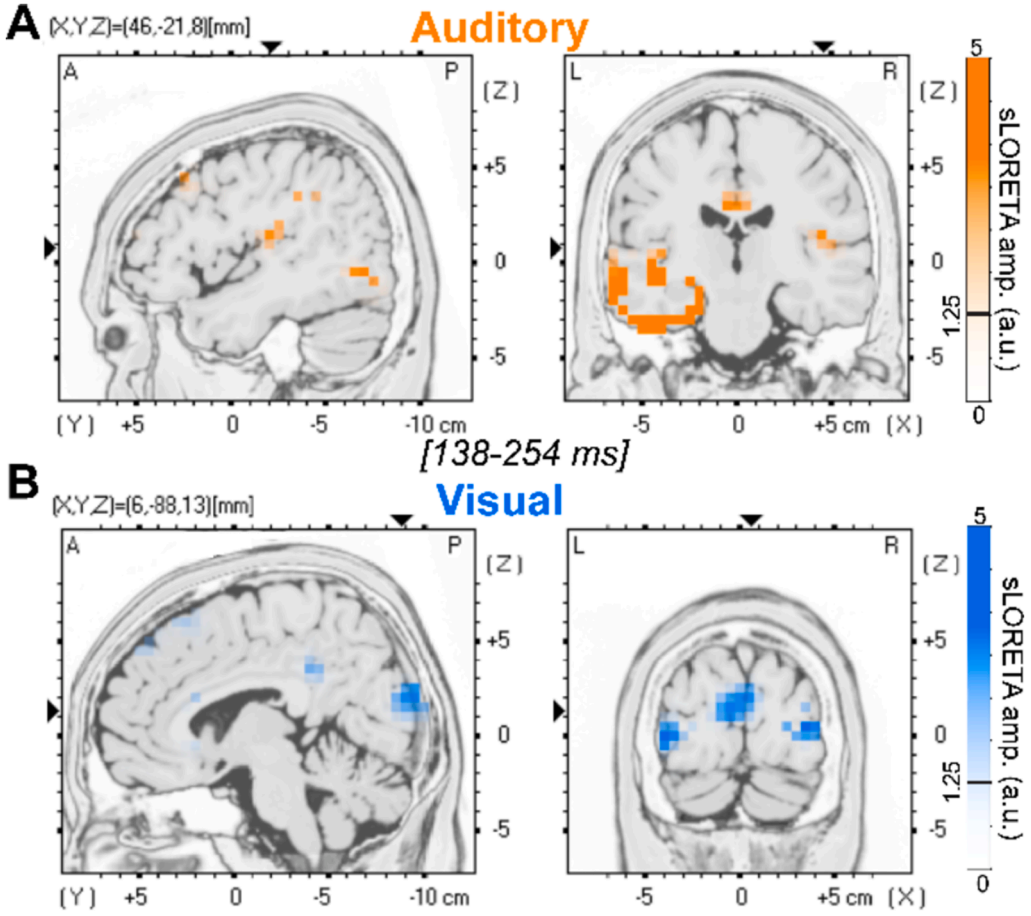

Fig. 6. Raw sLORETA source activations confirm modality-specific responses of auditory and visual cortex within 250 ms of stimulus onset. (A) AEP and (**B**) VEP source activations in the 138–254 ms time window that showed a significant main effect of modality (see †, Fig. 5A). Data are pooled across tokens within modality. Images are threshold masked at 1.25 sLORETA units for visualization.

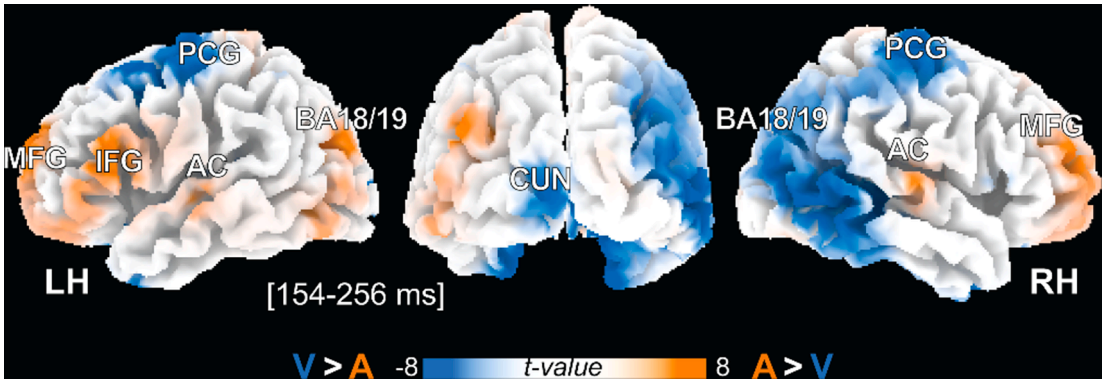

Fig. 7. Source responses reveal differential activation of auditory vs. visual category representation that depend on stimulus modality. sLORETA statistical maps contrasting A and V responses (t-stat, $p < 0.05$ masked, corrected) projected onto the Collins brain template (Collins et al., 1998). The contrast reflects the difference in degree of categorical coding between the auditory and visual modalities [i.e., $(A_{Tk1/5} - A_{Tk3}) - (V_{Tk1/5} - V_{Tk3})$]. Maps are shown in the 154–256 ms time window during the token x modality interaction (see *, Fig. 5A). Hot colors denote preferential coding of auditory categories; cool colors, preferential coding of visual categories. AC, auditory cortex; BA, Brodmann area; CUN, cuneus; MFG, middle frontal gyrus; IFG, inferior frontal gyrus; PCG, precentral gyrus; LH/RH, left/right hemisphere.

perceptually ambiguous than their auditory counterparts (Bidelman and Carter, 2023; Bidelman and Walker, 2017; Carter et al., 2022; Pisoni and Tash, 1974; Rizzi and Bidelman, 2024). Taken together, the overall sharper categorical pattern and faster overall RTs for auditory vs. visual tokens suggests stronger categorization for auditory phonemes compared to visual grapheme stimuli.

Paralleling the behavioral data, we found faster neural timing corresponded with faster perceptual categorization speeds. ERP responses also peaked ~ 200 ms earlier for auditory compared to visual stimuli. Moreover, MDS scaling of the EEG revealed responses self-organized in a categorial fashion such that tokens clustered within their respective modality beginning ~150–250 ms after stimulus onset. These data are

consistent with prior studies examining the neural differentiation of phonetic categories in the AEPs (Bidelman and Lee, 2015; Bidelman, et al., 2013; Chang, et al., 2010), where within-category tokens (e.g., Tk1/2) tend to cluster in closer proximity to one another but far from their across-category counterparts (e.g., Tk 4/5). Category clustering was also more prominent in the auditory than visual modality as indicated by a tighter convergence of within-category stimuli and farther separation across categories for phonemes. Collectively, our behavioral and neuroimaging data suggest the neural differentiation and subsequent grouping of tokens into categorical representations is more binary and robust in the auditory vs. visual modality for stimuli otherwise matched in linguistic identity.

Although our study only assessed uni-sensory audio/visual responses, our results extend a large body of work on multisensory integration in categorization. Integrating multiple cues is necessary in face-to-face communication in which visual articulatory information from a talker's face provides a critical complement to what was said. In audiovisual contexts, dynamic speech features in auditory and visual channels reflect discrete representations of phonetic-linguistic units (phonemes) and corresponding representations of mouth shapes (visemes) (Peelle and Sommers, 2015) that can interact to systematically influence the perceptual identity of speech objects themselves (Bidelman, et al., 2019; Massaro and Cohen, 1983; van Wassenhove et al., 2005). Electrophysiological studies have also shown that visual stimuli modulate auditory cortical responses in the auditory cortical fields (Kayser et al., 2008) and visual cues can increase the precision of category representations leading to a sharper perceptual division of the speech signal that aids its perception (Bidelman, et al., 2019). The activation of primary auditory cortex during lip reading further implies visual cues might influence perception even before speech sounds are categorized into their phonetic constituents (Bernstein and Liebenthal, 2014; Calvert et al., 1997).

Cross-modal interactions within sensory brain regions have also been observed in human neuromagnetic brain responses to auditory and visual stimuli (Raij et al., 2010). These studies reveal that while cross-sensory (auditory → visual) activity generally manifests later (~10–20 ms) than sensory-specific (auditory → auditory) activations, there is a stark asymmetry in the arrival of information between Heschl's gyrus and the Calcarine fissure. Auditory information is combined in visual cortex roughly 45 ms faster than the reverse direction of travel (i.e., visual → auditory) (Raij, et al., 2010) and auditory cues can bias and "override" normal perception of visual objects (Bidelman and Myers, 2020). Our data also broadly converge with other EEG studies examining audiovisual integration in speech (i.e., the McGurk effect) which suggests vision helps encode phonemic information within auditory cortex in a similar time window identified here (e.g., beginning at ~100 ms) (Abbott and Shahin, 2018; Shahin et al., 2018). Such dominance of auditory compared to visual information in multisensory studies might explain the larger and more extensive categorical organization we find for speech-sound phonemes compared to grapheme equivalents.

Phoneme and grapheme categorization recruit shared and segregated brain networks

At the scalp level, visual responses appeared to peak later than auditory responses (~200 vs. 400 ms; Fig. 3). At first glance, this might imply visual category information is somehow already present in extrastriate regions prior to perceptual coding. However, further scrutiny of waveform time courses (Fig. 3) and sources (Fig. 6) showed that first peak auditory and visual activations actually occurred in a similar timeframe (~200–250 ms) and in their respective temporal vs. occipital sensory cortex. Consequently, the later deflection in the VEPs we observe at ~300–400 ms may reflect extra post-perceptual processing that is evident in difficult categorization tasks (Bidelman, et al., 2020). This notion is consistent with our behavioral findings that visual grapheme were perceived less categorically than auditory phonemes (Fig. 2). However, there is EEG evidence that higher-order brain areas that support linguistic processes might indeed engage prior to sensory-perceptual coding. For example, inferior frontal language areas can show slightly earlier activity than auditory cortical regions during speech in noise tasks (Bidelman and Dexter, 2015). Relatedly, semantic responses can precede visual activation during semantic word judgments (Louwerse and Hutchinson, 2012). These findings are broadly consistent with "top-down" influences on perceptual processing (Pfingst and McKenzie, 2012) and could be achieved, among other means, by predictive coding or attentional biasing schemes that help anticipate behavioral output. Regardless, such findings underscore the notion that speech operations might not be entirely serial, but rather, the brain uses multiple routes for lexical access that are implemented in parallel processing channels (Hickok and Poeppel, 2007).

While TANOVA and MDS clustering of electrode responses revealed when A and V category representations emerge in the brain, the technique only operates on the global field power of the scalp data. As such, it did not allow us to identify which channels (or underlying brain areas) might drive modality-specific vs. modality-independent category representations. In this vein, source reconstruction allowed us to examine the brain regions supporting auditory vs. visual phonetic processing that is not possible from scalp-EEG alone (cf. Abbott and Shahin, 2018; Shahin, et al., 2018).

We had originally hypothesized sound phoneme and visual grapheme tokens representing the same linguistic identity might recruit common neural substrates (e.g., IFG), despite originating from different sensory modalities, which would have suggested a domain-general, isomorphic mapping of category representation. Instead, our hypothesis was only partially confirmed. Source analysis revealed distributed, but partially overlapping neural networks supporting phoneme vs. grapheme categorization. These findings are broadly consistent with previous functional connectivity studies that have identified a sparse but distributed brain network supporting phonetic categorization including areas of left linguistic (IFG), visual (cuneus/precuneus), and motor cortex (central gyrus) (Al-Fahad et al., 2020; Mahmud, et al., 2021). However, direct comparisons between continuum revealed category coding for speech sound phonemes was stronger than visual graphemes in bilateral auditory cortices, left IFG, and middle frontal gyrus (MFG). Engagement of auditory cortex in processing sound categories is consistent with prior work showing early auditory cortical areas typically associated with sensory-stimulus coding are highly sensitive to the category structure in speech (Bidelman and Lee, 2015; Bidelman and Walker, 2019; Carter and Bidelman, 2021; Chang, et al., 2010; Mankel et al., 2020; Rizzi and Bidelman, 2024). Similarly, left IFG has been implicated in phoneme category selectivity (Alho et al., 2016; Bidelman and Walker, 2019; Myers et al., 2009) and resolving ambiguity in the speech signal—as in cases of additive noise or lexical uncertainty (Carter and Bidelman, 2021; Luthra et al., 2019) (but see Hickok et al., 2011). We have also shown left MFG is recruited when listeners experience perceptual shifts in their hearing of speech categories dependent on top-down factors such as stimulus context or lexical biasing (Bidelman, et al., 2021; Carter, et al., 2022). Left IFG might also be involved in articulatory rehearsal during phonetic perception (Zatorre et al., 1992). Broadly speaking, the engagement of MFG and IFG in our auditory tasks is consistent with the notion that sound categorization recruits post-perceptual processing of the frontal lobes (Binder et al., 2004; Carter, et al., 2022; Myers, et al., 2009). However, we note these are relatively fast processes, engaging reciprocal auditory-frontal pathways by ~250 ms after sound enters the ear (see also Bidelman, et al., 2021; Mahmud, et al., 2021).

In contrast, we found visual categories recruited a network involving nodes in bilateral precentral gyrus (PCG), cuneus (CUN), and the right dorsal stream of the visual pathway including lateral occipital cortex and visual association areas (BA 18/19). These latter regions form the brain's canonical reading and visual word form areas (Selpien et al., 2015). Stronger engagement of auditory vs. visual cortex for phonemes

and graphemes, respectively, is perhaps expected and implies a unitary division for processing modality-specific sensory information. Indeed, as with primary auditory cortex (Bidelman and Lee, 2015; Chang, et al., 2010), category-specific information can be read out from early visual cortex (Vetter et al., 2014). Auditory activations have also been shown to predict phonological processing and rapid automatized naming whereas precuneus activations have been shown to predict reading and writing skills (Xu et al., 2018), respectively. Precentral engagement during visual grapheme tokens is also consistent with prior literature demonstrating engagement of primary motor cortex and supplementary motor areas during the perception of handwritten letters (Longcamp et al., 2011).

Surprisingly, we found auditory categories more strongly recruited portions of occipital cortex including extrastriate visual association areas (BA 18/19). These findings are interesting because they suggest nonretinal information is not only coded in the activity patterns of early visual cortex but that sounds might be processed in visual system in a form of speech imagery (Vetter et al., 2014). Indeed, some have argued that the left dominance for language even originates in extrastriate cortex (Selpien et al., 2015). However, we also found these effects depended on input modality and hemisphere. Whereas sound phonemes more strongly recruited *left* BA 18/19, letter graphemes recruited its homologue in *right* hemisphere. This implies a hemispheric asymmetry in the division of labor when processing acoustic vs. visual categories with the same linguistic-phonetic identity. While the basis of this asymmetry is not fully clear, it interesting to note that lateralization of ventral occipital responses varies with reading expertise (Seghier and Price, 2011). Neural activity is left lateralized for words in skilled readers but right lateralized in novice readers who have not yet learned to link print to sound (Maurer et al., 2006). Conceivably, lateralization might also vary according to individual differences in the related skill of letter-speech sound integration. Under this notion, the stark occipital lateralization we find for auditory (left) vs. visual (right) categorization may result from individual differences in orthographic decoding or auditory–visual matching. At the very least, our results suggest that extrastriate brain areas might perform an isomorphic, domain-general mapping for phonological categories in dorsal visual system. Such a computational hub would allow the brain to map linguistically-relevant sounds and visual linguistic objects alike into a common lexical (rather than purely auditory or visual) representation. Future studies employing audiovisual speech could test this possibility (e.g., Bidelman et al., 2019; Massaro and Cohen, 1983).

Author contributions

GMB designed the experiment, CP and AY collected the data, GMB analyzed the data, and all authors wrote the paper.

CRediT authorship contribution statement

Gavin M. Bidelman: Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Formal analysis, Data curation. **Ashleigh York:** Data curation. **Claire Pearson:** Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Work supported by the National Institutes of Health (NIH/NIDCD R01DC016267). The authors thank Gwyneth Lewis for assistance in data collection.

References

Abbott, N.T., Shahin, A.J., 2018. Cross-modal phonetic encoding facilitates the McGurk illusion and phonemic restoration. J. Neurophysiol. 120, 2988–3000.
Al-Fahad, R., Yeasin, M., Bidelman, G.M., 2020. Decoding of single-trial EEG reveals unique states of functional brain connectivity that drive rapid speech categorization decisions. J. Neural Eng. 17, 016045.
Alho, J., Green, B.M., May, P.J.C., Sams, M., Tiitinen, H., Rauschecker, J.P., Jääskeläinen, I.P., 2016. Early-latency categorical speech sound representations in the left inferior frontal gyrus. Neuroimage 129, 214–223.
Altmann, C.F., Uesaki, M., Ono, K., Matsuhashi, M., Mima, T., Fukuyama, H., 2014. Categorical speech perception during active discrimination of consonants and vowels. Neuropsychologia 64C, 13–23.
Bakdash, J.Z., Marusich, L.R., 2017. Repeated measures correlation. Front. Psychol. 8, 456.
Beale, J.M., Keil, F.C., 1995. Categorical effects in the perception of faces. Cognition 57, 217–239.
Bentin, S., 1992. Chapter 11 Phonological Awareness, Reading, and Reading Acquisition: A Survey and Appraisal of Current Knowledge. In: Frost, R., Katz, L. (Eds.), Advances in Psychology, vol. 94. North-Holland, pp. 193–210.
Bernstein, L.E., Liebenthal, E., 2014. Neural pathways for visual speech perception. Front. Neurosci. 8.
Bidelman, G.M., Myers, M.H., 2020. Frontal cortex selectively overrides auditory processing to bias perception for looming sonic motion. Brain Res. 1726, 146507.
Bidelman, G.M., Moreno, S., Alain, C., 2013. Tracing the emergence of categorical speech perception in the human auditory system. Neuroimage 79, 201–212.
Bidelman, G.M., Pousson, M., Dugas, C., Fehrenbach, A., 2018. Test-retest reliability of dual-recorded brainstem vs. cortical auditory evoked potentials to speech. J. Am. Acad. Audiol. 29, 164–174.
Bidelman, G.M., Sigley, L., Lewis, G., 2019. Acoustic noise and vision differentially warp speech categorization. J. Acoust. Soc. Am. 146, 60–70.
Bidelman, G.M., Walker, B.S., 2019. Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network. Neuroimage 201, 1–10.
Bidelman, G.M., Yellamsetty, A., 2017. Noise and pitch interact during the cortical segregation of concurrent speech. Hear. Res. 351, 34–44.
Bidelman, G.M., Bush, L.C., Boudreaux, A.M., 2020. Effects of noise on the behavioral and neural categorization of speech. Front. Neurosci. 14, 1–13.
Bidelman, G.M., Carter, J.A., 2023. Continuous dynamics in behavior reveal interactions between perceptual warping in categorization and speech-in-noise perception. Front. Neurosci. 17, 1–13.
Bidelman, G.M., Dexter, L., 2015. Bilinguals at the "cocktail party": dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners' speech-in-noise recognition deficits. Brain Lang. 143, 32–41.
Bidelman, G.M., Lee, C.-C., 2015. Effects of language experience and stimulus context on the neural organization and categorical perception of speech. Neuroimage 120, 191–200.
Bidelman, G.M., Walker, B., 2017. Attentional modulation and domain specificity underlying the neural organization of auditory categorical perception. Eur. J. Neurosci. 45, 690–699.
Bidelman, G.M., Pearson, C., Harrison, A., 2021. Lexical influences on categorical speech perception are driven by a temporoparietal circuit. J. Cogn. Neurosci. 33, 840–852.
Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., Ward, B.D., 2004. Neural correlates of sensory and decision processes in auditory object identification. Nat. Neurosci. 7, 295–301.
Burns, E.M., Campbell, S.L., 1994. Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: examples of categorical perception? J. Acoust. Soc. Am. 96, 2704–2719.
Burns, E.M., Ward, W.D., 1978. Categorical perception—phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. J. Acoust. Soc. Am. 63, 456–468.
Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P. K., Woodruff, P.W., Iversen, S.D., et al., 1997. Activation of auditory cortex during silent lipreading. Science 276, 593–596.
Carter, J.A., Bidelman, G.M., 2021. Auditory cortex is susceptible to lexical influence as revealed by informational vs. energetic masking of speech categorization. Brain Res. 1759, 147385.
Carter, J.A., Buder, E.H., Bidelman, G.M., 2022. Nonlinear dynamics in auditory cortical activity reveal the neural basis of perceptual warping in speech categorization. JASA Express Lett. 2, 045201.
Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., Knight, R.T., 2010. Categorical speech representation in human superior temporal gyrus. Nat. Neurosci. 13, 1428–1432.
Collins, D.L., Zijdenbos, A.P., Kollokian, V., et al., 1998. Design and construction of a realistic digital brain phantom. IEEE Trans. Med. Imaging 17, 463–468.
Ducati, A., Fava, E., Motti, E.D.F., 1988. Neuronal generators of the visual evoked potentials: intracerebral recording in awake humans. Electroencephalogr. Clin. Neurophysiol./Evoked Potentials Section 71, 89–99.
Ferraro, M., Foster, D.H., 1986. Discrete and continuous modes of curved-line discrimination controlled by effective stimulus duration. Spat. Vis. 1, 219–230.
Fonteneau, E., Davidoff, J., 2007. Neural correlates of colour categories. Neuroreport 18, 1323–1327.
Foster, D.H., 1983. Visual discrimination, categorical identification, and categorical rating in brief displays of curved lines: implications for discrete encoding processes. J. Exp. Psychol. Hum. Percept. Perform. 9, 785.

Foster, D.H., Ferraro, M., 1989. Visual gap and offset discrimination and its relation to categorical identification in brief line-element displays. J. Exp. Psychol. Hum. Percept. Perform. 15, 771.
Franklin, A., Drivonikou, G.V., Clifford, A., Kay, P., Regier, T., Davies, I.R., 2008. Lateralization of categorical perception of color changes with color term acquisition. Proc. Natl. Acad. Sci.:PNAS, 0809952105.
Godfrey, J.J., Syrdal-Lasky, A.K., Millay, K.K., Knox, C.M., 1981. Performance of dyslexic children on speech perception tests. J. Exp. Child Psychol.
Goldstone, R.L., Hendrickson, A.T., 2010. Categorical perception. Wiley Interdiscip. Rev. Cogn. Sci. 1, 69–78.
Habermann, M., Weusmann, D., Stein, M., Koenig, T., 2018. A student's guide to randomization statistics for multichannel event-related potentials using Ragu. Front. Neurosci. 12.
Harnad, S.R. (1987) Psychophysical and cognitive aspects of categorical perception: A critical overview. In: Categorical perception: The Groundwork of Cognition, vol. (Harnad SR, ed). New York: Cambridge University Press.
Hickok, G., Costanzo, M., Capasso, R., Miceli, G., 2011. The role of Broca's area in speech perception: evidence from aphasia revisited. Brain Lang. 119, 214–220.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.
Howard, D., Rosen, S., Broad, V., 1992. Major/minor triad identification and discrimination by musically trained and untrained listeners. Music Percept. Interdiscip. J. 10, 205–220.
Hu, L., Mouraux, A., Hu, Y., Iannetti, G.D., 2010. A novel approach for enhancing the signal-to-noise ratio and detecting automatically event-related potentials (ERPs) in single trials. Neuroimage 50, 99–111.
Jonathan, Morph (https://www.mathworks.com/matlabcentral/fileexchange/32683-morph), Retrieved Feb. 12, 2014, MATLAB Central File Exchange, 2011.
Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., TIrino, T., Banno, H., 2008. Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 3933–3936.
Kayser, C., Petkov, C.I., Logothetis, N.K., 2008. Visual modulation of neurons in auditory cortex. Cereb. Cortex 18, 1560–1574.
Klein, M.E., Zatorre, R.J., 2011. A role for the right superior temporal sulcus in categorical perception of musical chords. Neuropsychologia 49, 878–887.
Koenig, T., Melie-Garcia, L., 2010. A method to determine the presence of averaged event-related fields using randomization tests. Brain Topogr. 23, 233–242.
Koenig, T., Kottlow, M., Stein, M., Melie-García, L., 2011. Ragu: a free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. Comput. Intell. Neurosci. 2011, 938925.
Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. Psychol. Rev. 74, 431–461.
Liebenthal, E., Desai, R., Ellingson, M.M., Ramachandran, B., Desai, A., Binder, J.R., 2010. Specialization along the left superior temporal sulcus for auditory categorization. Cereb. Cortex 20, 2958–2970.
Locke, S., Kellar, L., 1973. Categorical perception in a non-linguistic mode. Cortex 9, 355–369.
Longcamp, M., Hlushchuk, Y., Hari, R., 2011. What differs in visual recognition of handwritten vs. printed letters? An fMRI study. Hum. Brain Mapp. 32, 1250–1259.
Louwerse, M., Hutchinson, S., 2012. Neurological evidence linguistic processes precede perceptual simulation in conceptual processing. Front. Psychol. 3.
Luthra, S., Guediche, S., Blumstein, S.E., Myers, E.B., 2019. Neural substrates of subphonemic variation and lexical competition in spoken word recognition. Language Cogn. Neurosci. 34, 151–169.
MacLean, J., Stirn, J., Sisson, A., Bidelman, G.M., 2024. Short- and long-term neuroplasticity interact during the perceptual learning of concurrent speech. Cereb. Cortex 34, 1–13.
Mahmud, M.S., Yeasin, M., Bidelman, G.M., 2021. Data-driven machine learning models for decoding speech categorization from evoked brain responses. J. Neural Eng. 18, 046012.
Mankel, K., Barber, J., Bidelman, G.M., 2020. Auditory categorical processing for speech is modulated by inherent musical listening skills. Neuroreport 31, 162–166.
Mankel, K., Shrestha, U., Tipirneni-Sajja, A., Bidelman, G.M., 2022. Functional plasticity coupled with structural predispositions in auditory cortex shape successful music category learning. Front. Neurosci. 16, 1–14.
Massaro, D.W., Cohen, M.M., 1983. Evaluation and integration of visual and auditory information in speech perception. J. Exp. Psychol. Hum. Percept. Perform. 9, 753–771.
Massaro, D.W., Hary, J.M., 1986. Addressing issues in letter recognition. Psychol. Res. 48, 123–132.
Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., Steinhausen, H.-C., Brandeis, D., 2006. Coarse neural tuning for print peaks when children learn to read. Neuroimage 33, 749–758.
McIntyre, M.C., Di Lollo, V., 1991. Categorical processing of visual stimuli in relation to geometrical, graphemic, or lexical context. Psychol. Res. 53, 142–148.
Michel, C.M., Murray, M.M., Lantz, G., Gonzalez, S., Spinelli, L., Grave de Peralta, R., 2004. EEG source imaging. Clin. Neurophysiol. 115, 2195–2222.
Miller, G.A., Gildea, P.M., 1987. How children learn words. Sci. Am. 257, 94–99.
Mody, M., Studdert-Kennedy, M., Brady, S., 1997. Speech perception deficits in poor readers: auditory processing or phonological coding? J. Exp. Child Psychol. 64, 199–231.
Moradi, S., Lidestam, B., Danielsson, H., Ng, E.H.N., Rönnberg, J., 2017. Visual cues contribute differentially to audiovisual perception of consonants and vowels in improving recognition and reducing cognitive demands in listeners with hearing impairment using hearing aids. J. Speech Lang. Hear. Res. 60, 2687–2703.
Murray, M.M., Brunet, D., Michel, C.M., 2008. Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr. 20, 249–264.
Myers, E.B., Blumstein, S.E., Walsh, E., Eliassen, J., 2009. Inferior frontal regions underlie the perception of phonetic category invariance. Psychol. Sci. 20, 895–903.
Nath, A.R., Beauchamp, M.S., 2012. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. Neuroimage 59, 781–787.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP measurements. Clin. Neurophysiol. 112, 713–719.
Pascual-Marqui, R.D., 2002. Standardized low resolution brain electromagnetic tomography (sLORETA). Methods Find. Exp. Clin. Pharmacol. 24D, 5–12.
Peelle, J.E., Sommers, M.S., 2015. Prediction and constraint in audiovisual speech perception. Cortex 68, 169–181.
Pfingst, K.A., McKenzie, D.N., 2012. The fusion of unattended duration representations as indexed by the mismatch negativity (MMN). Brain Res. 1435, 118–129.
Picton, T.W., Alain, C., Woods, D.L., John, M.S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., Trujillo, N.J., 1999. Intracerebral sources of human auditory-evoked potentials. Audiol. Neuro Otol. 4, 64–79.
Picton, T.W., van Roon, P., Armilio, M.L., Berg, P., Ille, N., Scherg, M., 2000. The correction of ocular artifacts: a topographic perspective. Clin. Neurophysiol. 111, 53–65.
Pisoni, D.B., 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Percept. Psychophys. 13, 253–260.
Pisoni, D.B., Luce, P.A., 1987. Acoustic-phonetic representations in word recognition. Cognition 25, 21–52.
Pisoni, D.B., Tash, J., 1974. Reaction times to comparisons within and across phonetic categories. Percept. Psychophys. 15, 285–290.
Preston, J.L., Molfese, P.J., Frost, S.J., Mencl, W.E., Fulbright, R.K., Hoeft, F., Landi, N., Shankweiler, D., et al., 2016. Print-speech convergence predicts future reading outcomes in early readers. Psychol. Sci. 27, 75–84.
Raij, T., Ahveninen, J., Lin, F.-H., Witzel, T., Jääskeläinen, I.P., Letham, B., Israeli, E., Sahyoun, C., et al., 2010. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. Eur. J. Neurosci. 31, 1772–1782.
Rizzi R, Bidelman GM (2024), Functional benefits of continuous vs. categorical listening strategies on the neural encoding and perception of noise-degraded speech. bioRxiv [preprint] doi: https://doi.org/10.1101/2024.05.15.594387.
Seghier, M.L., Price, C.J., 2011. Explaining left lateralization for words in the ventral occipitotemporal cortex. J. Neurosci. 31, 14745–14753.
Selpien, H., Siebert, C., Genc, E., Beste, C., Faustmann, P.M., Güntürkün, O., Ocklenburg, S., 2015. Left dominance for language perception starts in the extrastriate cortex: an ERP and sLORETA study. Behav. Brain Res. 291, 325–333.
Shahin, A.J., Backer, K.C., Rosenblum, L.D., Kerlin, J.R., 2018. Neural mechanisms underlying cross-modal phonetic encoding. J. Neurosci. 38, 1835–1849.
Siegel, J.A., Siegel, W., 1977. Categorical perception of tonal intervais: musicians can't tellsharp fromflat. Percept. Psychophys. 21, 399–407.
van Wassenhove, V., Grant, K.W., Poeppel, D., 2005. Visual speech speeds up the neural processing of auditory speech. PNAS 102, 1181–1186.
Vetter, P., Smith, F.W., Muckli, L., 2014. Decoding sound and imagery content in early visual cortex. Curr. Biol. 24, 1256–1262.
Werker, J.F., Tees, R.C., 1987. Speech perception in severely disabled and average reading children. Can. J. Psychol. 41, 48–61.
Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Wade, A.R., Boroditsky, L., 2007. Russian blues reveal effects of language on color discrimination. Proc. Natl. Acad. Sci. 104, 7780–7785.
Xu, W., Kolozsvari, O.B., Monto, S.P., Hämäläinen, J.A., 2018. Brain responses to letters and speech sounds and their correlations with cognitive skills related to reading in children. Front. Hum. Neurosci. 12, 1–17.
Yang, R.-X., Wang, W.S.Y., 2018. Categorical perception of Chinese characters by simplified and traditional Chinese readers. Read. Writ. 31, 1133–1154.
Yasuhara, M., Kuklinski, T.T., 1978. Category boundary effect for grapheme perception. Percept. Psychophys. 23, 97–104.
Zatorre, R.J., Halpern, A.R., 1979. Identification, discrimination, and selective adaptation of simultaneous musical intervals. Percept. Psychophys. 26, 384–395.
Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. Science (New York, N.Y.) 256, 846–849.