ARTICLE

Speech categorization is better described by induced rather than evoked neural activity

Md Sultan Mahmud,[1,a)] Mohammed Yeasin,[1] and Gavin M. Bidelman[2,b)]

[1]*Department of Electrical and Computer Engineering, University of Memphis, 3815 Central Avenue, Memphis, Tennessee 38152, USA*
[2]*School of Communication Sciences and Disorders, University of Memphis, 4055 North Park Loop, Memphis, Tennessee 38152, USA*

ABSTRACT:

Categorical perception (CP) describes how the human brain categorizes speech despite inherent acoustic variability. We examined neural correlates of CP in both evoked and induced electroencephalogram (EEG) activity to evaluate which mode best describes the process of speech categorization. Listeners labeled sounds from a vowel gradient while we recorded their EEGs. Using a source reconstructed EEG, we used band-specific evoked and induced neural activity to build parameter optimized support vector machine models to assess how well listeners' speech categorization could be decoded via whole-brain and hemisphere-specific responses. We found whole-brain evoked β-band activity decoded prototypical from ambiguous speech sounds with $\sim$70% accuracy. However, induced γ-band oscillations showed better decoding of speech categories with $\sim$95% accuracy compared to evoked β-band activity ($\sim$70% accuracy). Induced high frequency (γ-band) oscillations dominated CP decoding in the left hemisphere, whereas lower frequencies (θ-band) dominated the decoding in the right hemisphere. Moreover, feature selection identified 14 brain regions carrying induced activity and 22 regions of evoked activity that were most salient in describing category-level speech representations. Among the areas and neural regimes explored, induced γ-band modulations were most strongly associated with listeners' behavioral CP. The data suggest that the category-level organization of speech is dominated by relatively high frequency induced brain rhythms.
© 2021 Acoustical Society of America. https://doi.org/10.1121/10.0003572

I. INTRODUCTION

The human brain classifies diverse acoustic information into smaller, more meaningful groupings (Bidelman and Walker, 2017), a process known as categorical perception (CP). CP plays a critical role in auditory perception, speech acquisition, and language processing. Brain imaging studies have shown that neural responses elicited by prototypical speech sounds (i.e., those heard with a strong phonetic category) differentially engage Heschl's gyrus (HG) and the inferior frontal gyrus (IFG) compared to ambiguous speech (Bidelman *et al.*, 2013; Bidelman and Lee, 2015; Bidelman and Walker, 2017). Previous studies also demonstrate that the N1 and P2 waves of the event-related potentials (ERPs) are highly sensitive to speech perception and correlate with CP (Alain, 2007; Bidelman *et al.*, 2013; Mankel *et al.*, 2020). Our recent study (Mahmud *et al.*, 2020b) also demonstrated that different brain regions are associated with the encoding vs decision stages of processing while categorizing speech. These studies demonstrate that temporal dynamics of evoked activity provide a neural correlation of the different processes underlying speech categorization. However, ERP studies do not reveal how induced brain activity (so-called neural oscillations) might contribute to this process.

The electroencephalogram (EEG) can be divided into evoked (i.e., phase-locked) and induced (i.e., non-phase-locked) responses that vary in a frequency-specific manner (Shahin *et al.*, 2009). Evoked responses are largely related to the stimulus, whereas induced responses are additionally linked to different perceptual and cognitive processes that emerge during task engagement. These later brain oscillations (neural rhythms) play an important role in perceptual and cognitive processes and reflect different aspects of speech perception. For example, low frequency [e.g., θ (4–8 Hz)] bands are associated with syllable segmentation (Luo and Poeppel, 2012), whereas the α (9–13 Hz) band has been linked to attention (Klimesch, 2012) and speech intelligibility (Dimitrijevic *et al.*, 2017). Several studies report that listeners' speech categorization efficiency varies in accordance with their underlying induced and evoked neural activity (Bidelman *et al.*, 2013; Bidelman and Alain, 2015; Bidelman and Lee, 2015). For instance, Bidelman assessed correlations between ongoing neural activity (e.g., induced activity) and the slopes of listeners' identification functions, reflecting the strength of their CP (Bidelman, 2017). Listeners were slower and varied in their classification of more category-ambiguous speech sounds, which covaried with increases in induced γ activity (Bidelman, 2017). Changes in β (14–30 Hz) power are also strongly associated with listeners' speech identification skills (Bidelman, 2015). The β frequency band is linked with auditory template matching (Shahin *et al.*, 2009) between stimuli and internalized representations kept in memory (Bashivan *et al.*, 2014), whereas the higher γ frequency range (>30 Hz) is

associated with auditory object construction (Tallon-Baudry and Bertrand, 1999) and local network synchronization (Giraud and Poeppel, 2012; Haenschel *et al.*, 2000; Si *et al.*, 2017).

Studies also demonstrate hemispheric asymmetries in neural oscillations. During syllable processing, there is a dominance of γ frequency activity in the left hemisphere (LH) and θ frequency activity in the right hemisphere (RH) (Giraud *et al.*, 2007; Morillon *et al.*, 2012). Other studies show that during speech perception and production, lower frequency bands (3–6 Hz) better correlate with behavioral reaction times (RTs) than with higher frequencies (20–50 Hz; Yellamsetty and Bidelman, 2018). Moreover, the induced γ-band correlates with speech discrimination and perceptual computations during acoustic encoding (Ou and Law, 2018), further suggesting it reflects a neural representation of speech above and beyond evoked activity alone.

Still, given the high dimensionality of the EEG data, it remains unclear which frequency bands, brain regions, and "modes" of neural function (i.e., evoked vs induced signaling) are most conducive to describing the neurobiology of speech categorization. To this end, the recent application of machine learning (ML) to neuroscience data might prove useful in identifying the most salient features of brain activity that predict human behaviors. ML is an entirely data-driven approach that "decodes" neural data with minimal assumptions on the nature of exact representation or where those representations emerge. Germane to the current study, ML has been successfully applied to decode the speed of listeners' speech identification (Al-Fahad *et al.*, 2020) and related receptive language brain networks (Mahmud *et al.*, 2020b) from multichannel EEGs.

Departing from previous hypothesis-driven studies (Bidelman, 2017; Bidelman and Alain, 2015; Bidelman and Walker, 2017), the present work used a comprehensive data-driven approach [i.e., stability selection and support vector machine (SVM) classifier] to investigate the neural mechanisms of speech categorization using whole-brain electrophysiological data. Our goals were to evaluate which neural regime [i.e., evoked (phase-synchronized ERP) vs induced oscillations], frequency bands, and brain regions are most associated with CP using whole-brain activity via a data-driven approach. Based on prior work, we hypothesized that evoked and induced brain responses would both differentiate the degree to which speech sounds carry category-level information (i.e., prototypical vs ambiguous sounds from an acoustic-phonetic continuum). However, we predicted induced activity would best distinguish category-level speech representations, suggesting a dominance of endogenous brain rhythms in describing the neural underpinnings of CP.

II. MATERIALS & METHODS

A. Participants

Forty-eight young adults (15 males, 33 females; aged 18–33 years old) participated in the study (Bidelman *et al.*, 2020; Bidelman and Walker, 2017; Mankel *et al.*, 2020). All participants had normal hearing sensitivity [i.e., <25 dB HL (hearing level) between 250 and 8000 Hz] and no history of neurological disease. Listeners were right-handed, native English speakers, and had achieved a collegiate level of education. All participants were paid for their time and gave informed written consent in accordance with the Declaration of Helsinki and a protocol approved by the Institutional Review Board at the University of Memphis.

B. Stimuli and task

We used a synthetic five-step vowel token continuum to examine the most discriminating brain activity (i.e., evoked or induced activity) while categorizing prototypical vowel speech sounds from ambiguous speech (Bidelman *et al.*, 2013). Speech spectrograms are represented in Fig. 1. Each speech token was 100 ms, including 10 ms rise/fall to minimize the spectral splatter in the stimuli. Each speech token contained an identical voice fundamental frequency ($F0$), second ($F2$), and third formant ($F3$) frequencies ($F0$, 150 Hz; $F2$, 1090 Hz; and $F3$, 2350 Hz). The first formant ($F1$) was varied over five equidistant steps (430–730 Hz) to produce a perceptual continuum from /u/ to /a/.

Stimuli were delivered binaurally at an intensity of 83 dB SPL through earphones (ER 2; Etymotic Research, Elk Grove Village, IL, USA). Participants heard each token 150–200 times presented in random order. They were asked to label each sound in a binary identification task ("/u/" or "/a/") as fast and accurately as possible. Their responses and RTs were logged. The interstimulus interval (ISI) was jittered randomly between 400 and 600 ms with a 20 ms step.

C. EEG recordings and data pre-procedures

EEGs were recorded from 64 channels at standard 10–10 electrode locations on the scalp and digitized at 500 Hz using neuroscan amplifiers (SynAmps RT; Compumedics Neuroscan, Charlotte, NC, USA). Subsequent preprocessing was conducted in the Curry 7 (neuroimaging software suite; Compumedics Neuroscan, Charlotte, NC, USA), and customized routines were coded in MATLAB (MathWorks, Natick, MA, USA). Ocular artifacts (e.g., eye-blinks) were corrected in the continuous EEG using principal component analysis (PCA) and then filtered (1–100 Hz; notched filtered 60 Hz). Trials with voltage $\geq 125\,\mu V$ were discarded. Cleaned EEGs were then epoched into single trials (-200–800 ms, where $t = 0$ was the stimulus onset) and common average referenced. For details, see Bidelman *et al.* (2020) and Bidelman and Walker (2017).
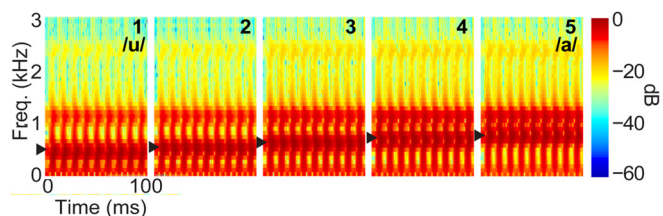

FIG. 1. (Color online) Speech stimuli. Acoustic spectrograms of the speech continuum from /u/ and /a/. Arrows, first formant frequency.

J. Acoust. Soc. Am. **149** (3), March 2021

Mahmud *et al.* 1645

D. EEG source localization

To disentangle the functional generators of CP-related EEG activity, we reconstructed the sources of the scalp recorded EEG by performing a distributed source analysis on single-trial data in Brainstorm software (Tadel *et al.*, 2011). We used a realistic boundary element head model (BEM) volume conductor and standard low-resolution brain electromagnetic tomography (sLORETA) as the inverse solution within Brainstorm (Tadel *et al.*, 2011). From each single-trial sLORETA volume, we extracted the time-courses within the 68 functional regions of interest (ROIs) across the LHs and RHs defined by the Desikan-Killiany (DK) atlas (Desikan *et al.*, 2006; LH, 34 ROIs; RH, 34 ROIs). Single-trial data were baseline corrected to the epoch's pre-stimulus interval (-200–0 ms).

Because we were interested in decoding prototypical (Tk1 and Tk5) from ambiguous speech (Tk3), we averaged responses to the end point tokens (hereafter referred to as Tk1/5) responses since they reflect prototypical vowel categories ("u" vs "a"). In contrast, Tk3 reflects an ambiguous category that listeners sometimes label as "u" or "a" (Bidelman *et al.*, 2020; Bidelman and Walker, 2017; Mankel *et al.*, 2020). To ensure an equal number of trials for prototypical and ambiguous stimuli, we considered 50% of the data from the merged Tk1/5 samples.

E. Time-frequency analysis

Time-frequency analysis was conducted via wavelet transform (Herrmann *et al.*, 2014). First, we computed the ERP using bootstrapping by randomly averaged over 100 trials with the replacement 30 times (Al-Fahad *et al.*, 2020) for each stimulus condition (e.g., Tk1/5 and Tk3) per subject and source ROI (e.g., 68 ROIs). We then applied the Morlet wavelet transform to each ROI average data (i.e., ERP) with time steps of 2 ms and an increment step frequency of 1 Hz from low to high frequency (e.g., 1–100 Hz) across the epoch, which provided only evoked frequency-specific activity (i.e., time- and phase-locked to stimulus onset). For computing induced activity, we performed a similar Morlet wavelet transform on a single-trial basis for each ROI and then computed the absolute value of each trial spectrogram. Next, we averaged the resulting time-frequency decompositions (Herrmann *et al.*, 2014), resulting in a spectral representation that contains the total activity. To isolate induced responses, we subtracted the evoked activity from the total activity (Herrmann *et al.*, 2014). We then extracted the different frequency band signals from the evoked and induced activity time-frequency maps for each brain region (e.g., 68 ROIs). Example evoked and induced time-frequency maps from the primary auditory cortex [i.e., transverse temporal (TRANs)] are represented in Fig. 2. We did not separate early vs late windows in this study as we have previously shown induced activity during speech categorization tasks is largely independent of motor responses (Bidelman, 2015).

Spectral features of different bands (θ, α, β, and γ) were quantified as the mean power over the full epoch. We concatenated four frequency bands that resulted in 4 $\times$ 68 $=$ 272 features for each response type (e.g., evoked vs induced) per speech condition (Tk1/5 vs Tk3). We conducted a paired t-test between evoked and induced feature vectors [e.g., concatenating all frequency band features of each ROI and stimulus type (Tk1/5 vs Tk3)] and found statistical significance [$t(783\,359) = 1212.53$, $p < 0.001$] between the two brain modes. We also conducted a one-way analysis of variance (ANOVA) to test for a band effect (θ-, α-, β-, and γ-band activity) within each brain regime. Band modulations were evident in both induced [$F(4,2876) = 247\,499.16$, $p < 0.001$] and evoked [$F(4,2876) = 108\,336.47$, $p < 0.001$] activities. To assess which regime (evoked vs induced) and oscillatory band (θ, α, β, and γ) is more important to speech categorization, we next used ML classifiers to decode the data. We separately (i.e., evoked and induced) submitted the individual frequency bands to the SVM and k-nearest neighbor (KNN) classifiers and all concatenated features (e.g., θ, α, β, and γ) to stability selection to investigate which frequency bands and brain regions decode prototypical vowels (e.g., Tk1/5) from the ambiguous (Tk3) vowel. The features were z-scored prior to the SVM and KNN to normalize them to a common range.

F. SVM classification

Our main goal was to decode prototypical vs ambiguous speech rather than investigate individual differences in speech categorization *per se*. Thus, all trials across listeners were used in the analysis without regard for individuals to decode the speech stimulus categories (Tk1/5 vs Tk3) from the EEGs at the group level ($N = 48$ participants). We used the parameter optimized SVM that yields a better classification performance with small sample size data (Bidelman *et al.*, 2019; Durgesh and Lekha, 2010; Mahmud *et al.*, 2020b). We chose SVM since our recent study (Mahmud *et al.*, 2020a) demonstrated this classifier yields slightly better performance as compared to KNN and AdaBoost for binary classification using EEG spectral features. (Note, we report decoding results using KNN classifiers in the Appendix as a confirmatory analysis to corroborate the main SVM findings.) The tunable parameters (e.g., kernel, C, γ) in the SVM model greatly affect the classification performance. We randomly split the data into training and test sets at 80% and 20%, respectively. During the training phase (i.e., using 80% data), we conducted a grid search approach with fivefold cross-validation, kernels $=$ "Radial basis function (RBF)" fine-tuned the C, and γ parameters to find the optimal values so that the classifier can accurately distinguish prototypical vs ambiguous speech (Tk1/5 vs Tk3) in the test data that models have never seen. Once the models were trained, we selected the best model with the optimal value of C and γ and predicted the unseen test data (by providing the attributes but no class labels). Classification performance metrics (accuracy, F1-score, precision, and recall) were calculated using standard formulas. The optimal values of C and γ for different analysis scenarios are given in the Appendix.

1646   J. Acoust. Soc. Am. **149** (3), March 2021
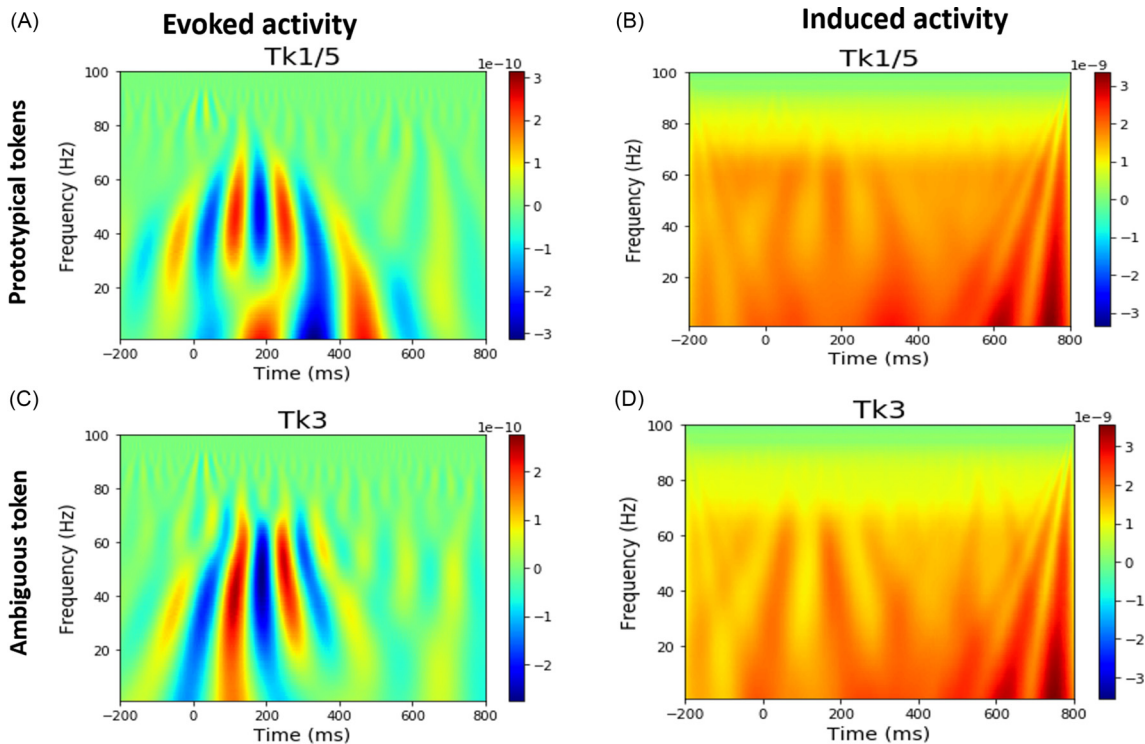
Mahmud *et al.*

FIG. 2. (Color online) Grand average neural oscillatory responses to prototypical vowel [e.g., Tk1/5 and ambiguous speech token (Tk3)]. [(A),(C)] Evoked activity for prototypical vs ambiguous tokens. [(B),(D)] Induced activity for prototypical vs ambiguous tokens. Responses are from primary auditory cortex (PAC; lTRANS, left transverse temporal gyrus).

G. Stability selection to identify critical brain regions of CP

Our data comprised a large number [68 ROIs × 4 bands (θ, α, β, and γ) = 272 features] of spectral measurements for each stimulus condition of interest (e.g., Tk1/5 vs Tk3) per ROI and brain regime (e.g., evoked and induced activities). We aimed to select a limited set of the most salient discriminating features for evoked and induced regimes via stability selection (Meinshausen and Bühlmann, 2010; see details in the Appendix).

During the stability selection implementation, we considered a sample fraction = 0.75, the number of resamples = 1000, and tolerance = 0.01 (Meinshausen and Bühlmann, 2010). In the Lasso (least absolute shrinkage and selection operator) algorithm, the feature scores were scaled between 0 and 1, where 0 is the lowest score (i.e., irrelevant feature) and 1 is the highest score (i.e., most salient or stable feature). We estimated the regularization parameter from the data using the least angle regression (LARs) algorithm. Over 1000 iterations, the randomized Lasso algorithm provided the overall feature scores (0–1) based on the number of times a variable was selected. We ranked stability scores to identify the most important, consistent, stable, and invariant features that could decode speech categories via the EEG. We submitted these ranked features and corresponding class labels to a SVM classifier with different stability thresholds and observed the model performance. We used this process separately for induced and evoked brain activities.

III. RESULTS

A. Behavioral results

Listeners' behavioral identification (%) functions and RTs (ms) for speech token categorizations are illustrated in Figs. 3(B) and 3(C), respectively. Responses abruptly shifted in speech identity (/u/ vs /a/) near the midpoint of the continuum, reflecting a change in the perceived category. The behavioral speed of speech labeling (e.g., RT) was computed from listeners' median response latencies for a given condition across all trials. RTs outside of 250–2500 ms were deemed outliers and excluded from further analysis (Bidelman et al., 2013; Bidelman and Walker, 2017). For each continuum, individual identification scores were fit with a two-parameter sigmoid function; $P = 1/[1 + e^{-\beta 1(x - \beta 0)}]$ where P is the proportion of the trial identification as a function of a given vowel, x is the step number along the stimulus continuum, and $\beta 0$ and $\beta 1$ are the location and slope of the logistic fit estimated using the nonlinear least squares regression, respectively (Bidelman and Walker, 2017). The slopes of the sigmoidal psychometric functions reflect the strength of CP [Fig. 3(A)].

B. Decoding categorical neural responses using band frequency features and SVM

We investigated the decoding of prototypical vowels from the ambiguous vowel (i.e., category-level representations) using the SVM neural classifier on the whole-brain (all 68 ROIs) and individual hemisphere (LH and RH) data
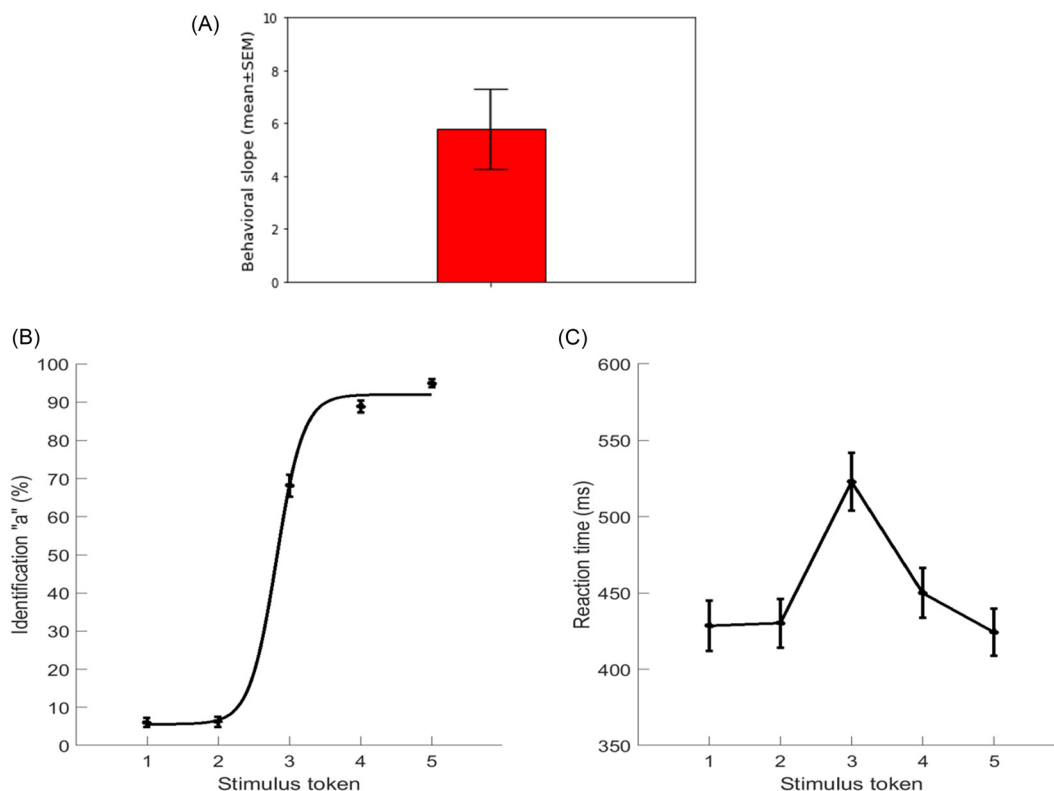
J. Acoust. Soc. Am. **149** (3), March 2021

Mahmud *et al.* 1647

FIG. 3. (Color online) Behavioral results. (A) Behavioral slope. (B) Psychometric functions showing % "a" identification of each token. Listeners' perception abruptly shifts near the continuum midpoint, reflecting a flip in the perceived phonetic category (i.e., "u" to "a"). (C) RTs for identifying each token. RTs are faster for prototype tokens (i.e., Tk1/5) and slow when categorizing ambiguous tokens at the continuum's midpoint (i.e., Tk3). Errorbars = ±1 s.e.m. (standard error of the mean).

separately for induced vs evoked activity (for comparable KNN classifier results see Appendix). The best model performance on the test dataset is reported in Fig. 4 and Table I. The mean accuracy and error variance of the fivefold cross-validation accuracy is reported in the Appendix (Fig. 8).

Using whole-brain *evoked* θ, α, β, and γ frequency responses, speech stimuli (e.g., Tk1/5 vs Tk 3) were correctly distinguished at 66%–69% accuracy. Among all of the evoked frequency bands, β-band was optimal to decode the speech categories (69.61% accuracy). The LH data revealed that θ-, α-, β-, and γ-bands decoded speech stimuli at accuracies between ~63% and 65%, whereas decoding from the RH was slightly poorer at 57%–62%.

Using whole-brain *induced* θ, α, β, and γ frequency responses, speech stimuli were decodable at accuracies of 89%–95%. Among all of the induced frequency bands, the γ band showed the best speech segregation (94.9% accuracy). Hemisphere-specific data again showed lower accuracy. LH oscillations decoded speech categories at 76%–87% accuracy, whereas the RH yielded 80%–84% accuracy.

C. Decoding brain regions associated with CP (evoked vs induced)

We separately applied the stability selection (Meinshausen and Bühlmann, 2010) to induced and evoked activity features to identify the most critical brain areas (e.g., ROIs) that have been linked with speech categorization.

Spectral features of brain ROIs were considered stable if the speech decoding accuracy was >70%. The effects of the stability scores on speech sound classification are represented in Fig. 5. Each bin of the histogram illustrates the number of features in a range of stability scores. In this work, the number of features (labeled in Fig. 5) represents the neural activity of different frequency bands, and the unique brain regions (labeled as ROIs in Fig. 5) represent the distinct functional brain regions of the DK atlas. The semi bell-shaped solid black and dotted red lines demonstrate the classifier accuracy and area under the curve (AUC), respectively. We submitted the neural features identified at different stability thresholds to the SVMs. This allowed us to determine whether the collection of neural measures identified via ML were relevant to classifying speech sound categorization.

For induced responses, most features (60%) yielded stability scores of 0–0.1, meaning 163/272 (60%) were selected less than 10% of the time out of 1000 iterations from 68 ROIs. A stability score of 0.2 selected 47/86 (32%) of the features from 47 ROIs that could decode speech categories at 96.9% accuracy. The decoding performance decreased with an increasing stability score (i.e., more conservative variable/brain ROI selections), resulting in a reduced feature set that retained only the most meaningful features distinguishing speech categories from a few ROIs. For instance, corresponding to the stability threshold 0.5, 25 (10%) features were selected from 21 brain ROIs that yielded the speech

1648   J. Acoust. Soc. Am. **149** (3), March 2021
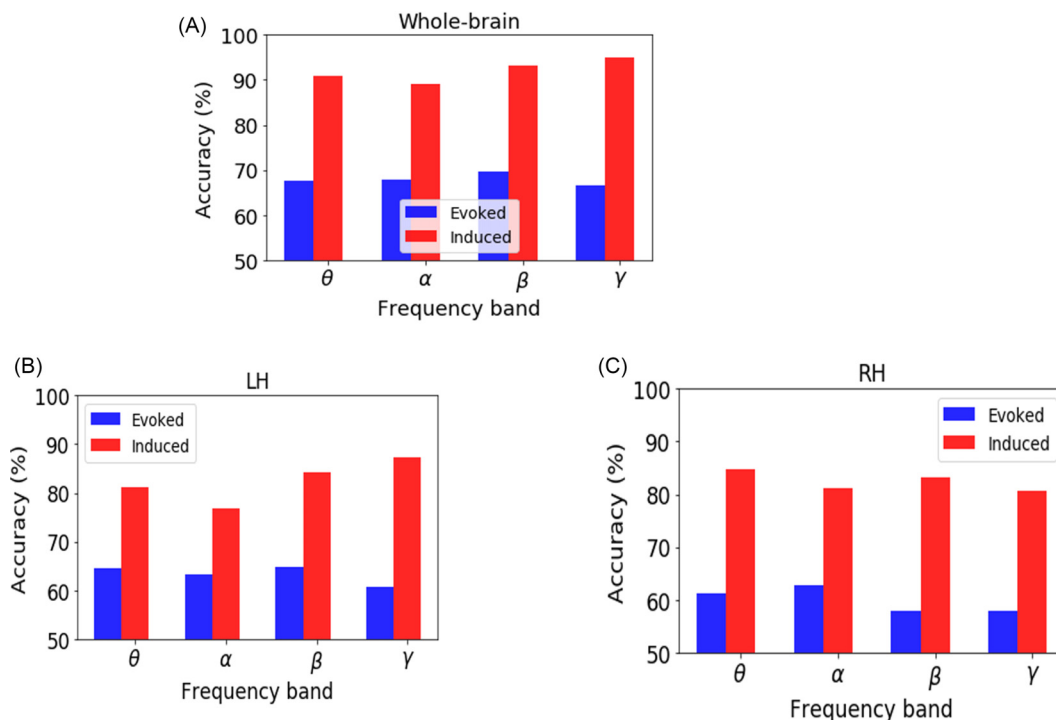
Mahmud *et al.*

FIG. 4. (Color online) Decoding categorical neural encoding using different frequency band features of the source-level EEG. The SVM results classifying prototypical (Tk1/5) vs ambiguous (Tk 3) speech sounds. (A) Whole-brain data (e.g., 68 ROIs), (B) LH (e.g., 34 ROIs), and (C) RH (e.g., 34 ROIs). Chance level = 50%.

categorization 92.7% accurately. However, corresponding to the stability threshold 0.8, only two features were selected from two brain ROIs that decoded the CP at 60.6%, still greater than the chance level (i.e., 50%). Performance improved by ~10% (86.5%) when the stability score was changed from 0.7 (selected brain ROIs is 9) to 0.6 (selected brain ROIs is 14). A BrainO visualization (Moinuddin *et al.*, 2019) of these brain ROIs is shown in Fig. 6.

Using evoked activity, the maximum decoding accuracy was 78.0% at a 0.1 stability threshold. Here, 43% of the features produced a stability score between 0.0 and 0.1. These 118 (43%) features are not informative because they decreased the model's accuracy to properly categorize speech. Corresponding to the stability scores 0.9, only eight features were selected from the six brain ROIs, which decoded speech at 65.8% accuracy. At a stability score 0.6, 29 (1%) features were selected from 22 brain ROIs, corresponding to an accuracy performance of 71.4%.

Our goal is to build an interpretable model that can describe speech categorization with reasonable accuracy using the smallest number of brain ROIs/features. Usually, the knee point is a location along the stability curve which balances model complexity (i.e., feature count) and decoding accuracy. Thus, 0.6 might be considered an optimal stability score (i.e., the knee point of a function in Fig. 5) as it decoded speech well above the change ($>$70%) with a minimal (and, therefore, more interpretable) feature set for both induced and evoked activity. Brain ROIs corresponding to the optimal stability score (0.6) are depicted in Fig. 7 and Tables II and III for both evoked and induced activities.

D. Brain-behavior relationships

To examine the behavioral relevance of the brain ROIs identified in the stability selection, we conducted the multivariate weighted least square (WLS) analysis regression

TABLE I. Performance metrics of the SVM classifier for decoding prototypical vs ambiguous vowels (i.e., categorical neural responses) using whole-brain data.

Neural activity	Frequency band	Accuracy (%)	AUC (%)	Precision (%)	Recall (%)	F1-score (%)
Evoked	θ	67.53	67.59	68.00	68.00	67.00
	α	67.88	67.92	68.00	68.00	67.00
	β	69.61	69.64	68.00	68.00	68.00
	γ	66.66	66.65	67.00	67.00	67.00
Induced	θ	90.97	90.98	91.00	91.00	91.00
	α	89.06	89.05	89.00	89.00	89.00
	β	93.23	93.20	93.00	93.00	93.00
	γ	94.96	94.96	95.00	95.00	95.00

J. Acoust. Soc. Am. **149** (3), March 2021
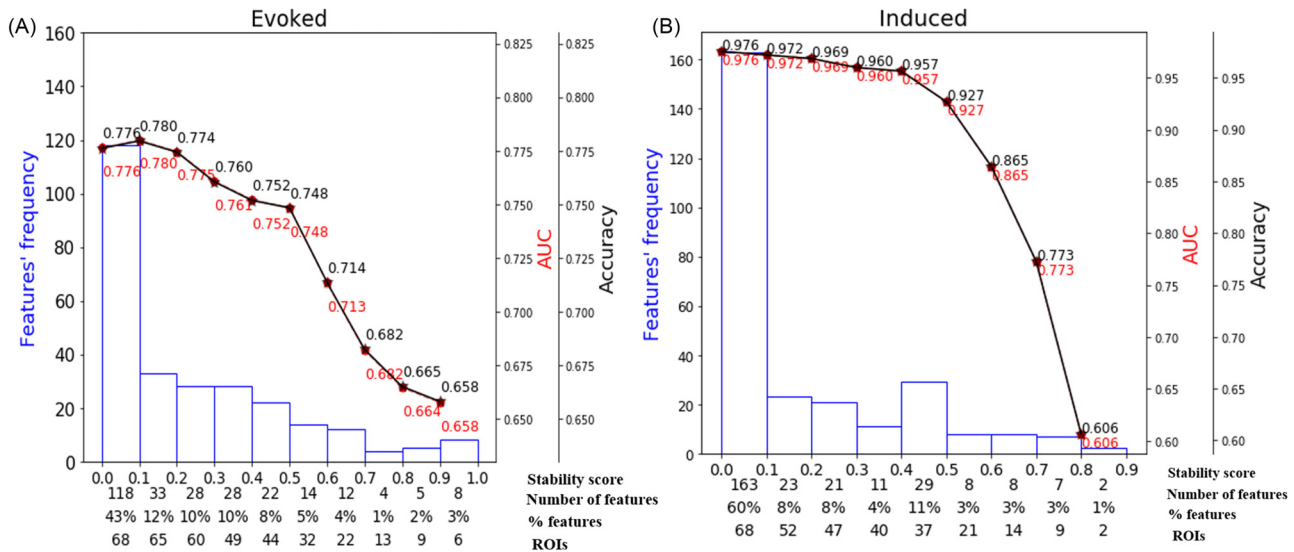
Mahmud *et al.* 1649

FIG. 5. (Color online) Effect of stability score threshold on model performance during (A) evoked activity and (B) induced activity during the CP task. The bottom of the x axis has four labels. *Stability score* represents the stability score range of each bin (scores range, 0–1); *number of features* represents the number of selected features under each bin; *% features* represents the corresponding percentage of selected features; and *ROIs* represents the number of cumulative unique brain regions up to the lower boundary of the bin.

(Ruppert and Wand, 1994). We conducted the WLS between the individual frequency band features (i.e., evoked and induced) and the slopes of the behavioral identification functions [i.e., Fig. 3(A)], which indexes the strength of listeners' CP. The WLS regression for *induced* activity is shown in Table II and the WLS regression for *evoked* activity is shown

in Table III. From the induced data, we found that γ frequency activity from six ROIs predicted behavior best among all other frequencies, $R^2 = 0.915$, $p < 0.0001$. Remarkably, only two brain regions [including the primary auditory cortex (PAC) and rostral anterior cingulate L] of the β-band frequency could predict behavioral slopes
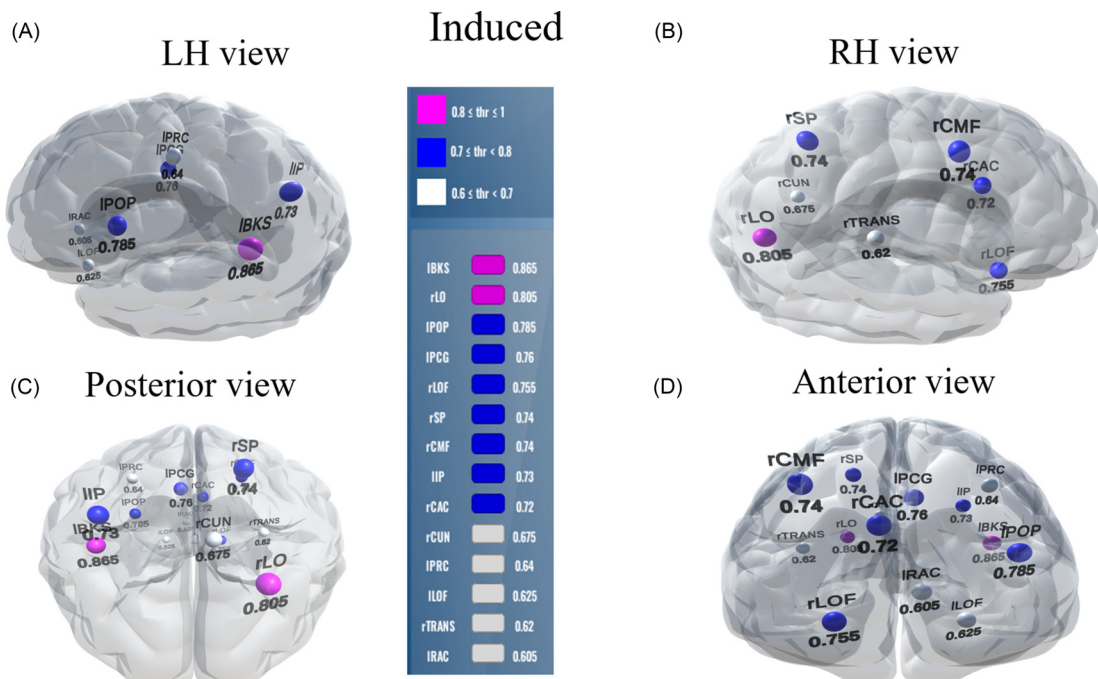

FIG. 6. (Color online) Stable (most consistent) neural network decoding using induced activity. Visualization of brain ROIs corresponding to ≥0.60 stability threshold (14 top selected ROIs), which show categorical organization (e.g., Tk1/5 ≠ Tk3) at 86.5%. (A) LH view (B) RH view, (C) posterior view, and (D) anterior view. Color legend demarcations show high (pink), moderate (blue), and low (white) stability scores. l/r, left/right; BKS, bankssts; LO, lateral occipital; POP, pars opercularis; PCG, posterior cingulate; LOF, lateral orbitofrontal; SP, superior parietal; CMF, caudal middle frontal; IP, inferior parietal; CAC, caudal anterior cingulate; CUN, cuneus; PRC, precentral; TRANS, transverse temporal; RAC, rostral anterior cingulate.
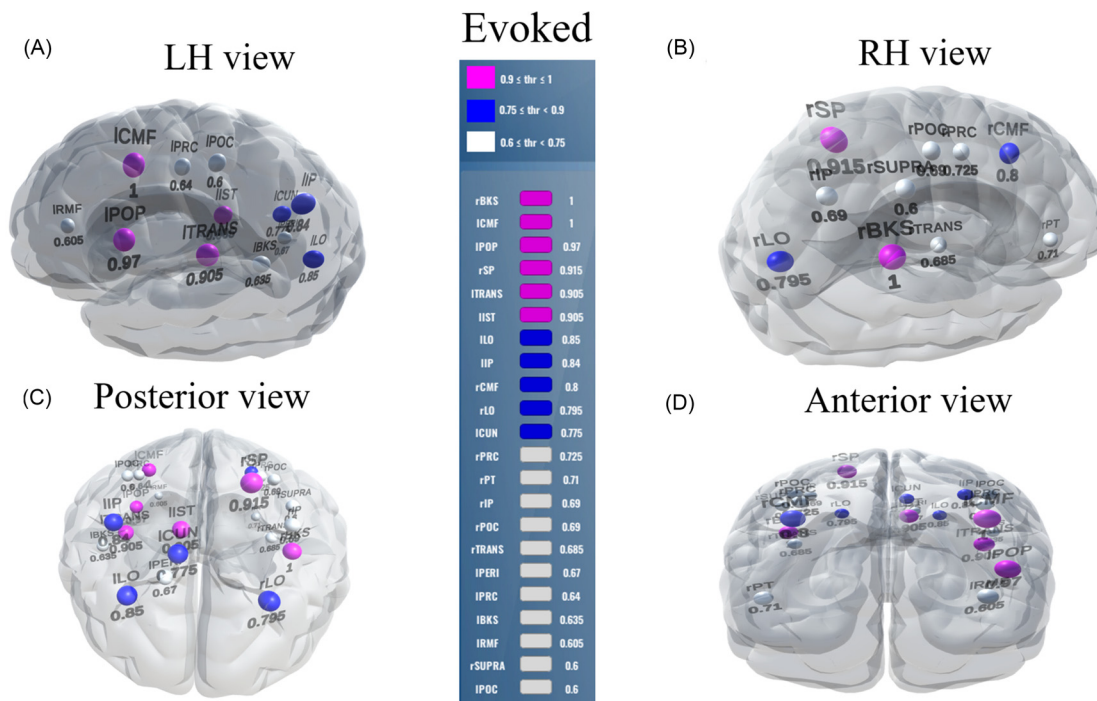
FIG. 7. (Color online) Stable (most consistent) neural network decoded using evoked activity. Visualization of brain ROIs corresponding to ≥0.60 stability threshold (22 top selected ROIs), which decode Tk1/5 from Tk3 at 71.4%. Otherwise, as in Fig. 6. BKS, bankssts; CMF, caudal middle frontal; POP, pars opercularis; SP, superior parietal; TRANS, transverse temporal; IST, isthmus cingulate; LO, lateral occipital; IP, inferior parietal; CUN, cuneus; PRC, pre-central; PT, pars triangularis; POC, postcentral; PERI, Pericalcarine; SUPRA, supra marginal.

$(R^2 = 0.876, p < 0.00001)$. Except in the α frequency band, evoked activity was poorer at predicting the behavioral CP.

IV. DISCUSSION

A. Speech categorization from evoked and induced activity

The present study aimed to examine which modes of brain activity and frequency bands of the EEG best decode speech categories and the process of categorization. Our results demonstrate that at the whole-brain level, evoked β-

band oscillations robustly code (~70% accuracy) the category structure of speech sounds. However, the induced γ-band showed better performance, classifying speech categories at ~95% accuracy, better than all of the other induced frequency bands. Our data are consistent with notions that higher frequency bands are associated with speech identification accuracy and carry information related to acoustic features and quality of speech representations (Yellamsetty and Bidelman, 2018). Our results also corroborate previous studies that suggest that higher frequency channels of the EEG (β, γ) reflect auditory-perceptual object construction

TABLE II. Brain-behavior relations of 14 brain ROIs in different frequency bands and behavioral predictions from the *induced* activity at a stability threshold ≥0.6, which yielded an accuracy of 86.5%.

Frequency band and combined R^2	ROI name	ROI abbreviation	Coefficient	p-value	Stability score
$\theta, R^2 = 0.807, p < 0.00001$	Pars opercularis L	lPOP	0.974	0.013	0.785[a]
	Posterior cingulate L	lPCG	−1.759	0.001	0.760
	Caudal anterior cingulate R	rCAC	3.163	0.001	0.740
$\alpha, R^2 = 0.746, p < 0.00001$	Bankssts L	lBKS	0.645	0.249	0.865
	Inferior parietal L	lIP	1.594	0.035	0.730
	Precentral L	lPRC	1.006	0.117	0.640
$\beta, R^2 = 0.876, p < 0.00001$	TRANs R	rTRANS	0.267	0.584	0.620
	Rostral anterior cingulate L	lRAC	3.004	0.001	0.605
$\gamma, R^2 = 0.915, p < 0.00001$	Lateral occipital R	rLO	0.768	0.804	0.805
	Lateral orbitofrontal R	rLOF	5.092	0.001	0.755
	Superior parietal R	rSP	16.472	0.004	0.740
	Caudal middle frontal R	rCMF	−3.243	0.188	0.740
	cuneus R	rCUN	−1.743	0.701	0.675
	Lateral orbitofrontal L	lLOF	0.709	0.553	0.625

[a]A score of 0.785, for example, means that out of 1000 iterations, the ERP feature of this ROI was selected 785 times by stability selection.

J. Acoust. Soc. Am. **149** (3), March 2021

Mahmud *et al.*   1651

TABLE III. Brain-behavior relations of 22 brain ROIs in different frequency bands and behavioral slope predictions from the *evoked* activity at a stability threshold ≥ 0.6, which yielded an accuracy of 71.4%.

Frequency band and combined R^2	ROI name	ROI abbreviation	Coefficient	p-value	Stability score
$\theta, R^2 = 0.349, p < 0.0184$	Caudal middle frontal L	lCMF	−93.646	0.575	1
	Superior parietal R	rSP	−89.350	0.603	0.915
	Isthmus cingulate L	lIST	−46.527	0.671	0.905
	Lateral occipital L	lLO	−190.348	0.149	0.850
	Pars triangularis R	rPT	137.923	0.160	0.710
	Post central R	rPOC	180.073	0.015	0.690
	Rostral middle frontal L	lRMF	−69.220	0.326	0.605
	Post central L	lPOC	152.979	0.238	0.600
$\alpha, R^2 = 0.863, p < 0.00001$	Bankssts R	rBKS	−64.139	0.775	1
	TRANs L	lTRANS	62.583	0.729	0.905
	Inferior parietal L	lIP	986.399	0.027	0.840
	Caudal middle frontal R	rCMF	−254.140	0.338	0.800
	Inferior parietal R	rIP	−707.049	0.053	0.690
	Pericalcarine L	lPERI	−278.456	0.319	0.670
	Precentral L	lPRC	163.234	0.368	0.640
	Bankssts L	lBKS	947.797	0.0001	0.635
	Supra marginal R	rSUPRA	−466.985	0.107	0.600
$\beta, R^2 = 0.198, p < 0.0184$	Pars opercularis L	lPOP	475.923	0.240	0.970
	Lateral occipital R	rLO	−1119.991	0.157	0.795
	Precentral R	rPRC	1485.600	0.008	0.725
$\gamma, R^2 = 0.604, p < 0.00001$	Cuneus L	lCUN	−3730.107	0.160	0.775
	TRANs R	rTRANS	262.544	0.0001	0.685

(Tallon-Baudry and Bertrand, 1999) and how well listeners map vowel sounds to category labels (Bidelman, 2015, 2017).

Analysis by hemisphere showed that induced γ activity was dominant in the LH, whereas lower frequency bands (e.g., θ) were more dominant in the RH. These findings support the asymmetric engagement of frequency bands during syllable processing (Giraud *et al.*, 2007; Morillon *et al.*, 2012) and the lower frequency bands in RH dominance in inhibitory and attentional control (top-down processing during complex tasks; Garavan *et al.*, 1999; Price *et al.*, 2019). Our results are consistent with the idea that cortical theta and gamma frequency bands play a key role in speech encoding (Hyafil *et al.*, 2015). They also show that the ML model was able to decode acoustic-phonetic information (i.e., speech categories) in the LH (using induced high frequency) and in the RH (using low frequencies).

B. Brain networks involved in speech categorization

ML (stability selection coupled with SVM) further identified the most stable, relevant, and invariant brain regions that associate with speech categorization. Stability selection identified 22 and 14 critical brain ROIs using evoked and induced activity, respectively. Our results show that induced activity better characterizes speech categorization using less neural resources (i.e., fewer brain regions) as compared to evoked activity. Eight brain ROIs (e.g., bankssts L, lateral occipital R, pars opercularis L, superior parietal R, caudal middle frontal R, inferior parietal L, precentral L, TRANs R) are common in evoked and induced regimes. These eight areas included the primary auditory cortex (TRANs R), Brocas's area (pars opercularis L), and the motor area (precentral L), which are critical

to speech-language processing. Superior parietal and inferior parietal areas have been associated with auditory, phoneme, and sound categorizations in particularly ambiguous contexts (Dufor *et al.*, 2007; Feng *et al.*, 2018). For the nonoverlapping areas in the induced activity, orbitofrontal is associated with speech comprehension and rostral anterior cingulate is associated with speech control (Sabri *et al.*, 2008). Surprisingly, out of the identified 14 brain ROIs, 3 ROIs are in θ, 3 are in α, 2 are in β, and 6 are in the γ-band. Notably, we found that a greater number of brain regions were recruited in the γ-frequency band. This result is consistent with the notion that high frequency oscillations play a role in network synchronization and widespread construction of perceptual objects related to abstract speech categories (Giraud and Poeppel, 2012; Haenschel *et al.*, 2000; Si *et al.*, 2017; Tallon-Baudry and Bertrand, 1999). Indeed, γ-band activity in only six ROIs was the best predictor of the listeners' behavioral speech categorization. Interestingly, nine ROIs of evoked α-band activity were able to predict behavioral slopes better than the induced α-band activity. This result supports notions that the α frequency band is associated with attention (Klimesch, 2012) and speech intelligibility (Dimitrijevic *et al.*, 2017).

A main advantage of our data-driven approach is that it identifies the frequency bands and brain regions that are best linked to speech categorization behaviors from among the many thousands of features measurable from the whole-brain EEG. It is a complement to conventional hypothesis-driven approaches (Bidelman, 2015; Bidelman and Alain, 2015; Bidelman and Walker, 2019) but is perhaps more hands off in that it requires fewer assumptions about the underlying brain mechanisms supporting speech perception.

1652   J. Acoust. Soc. Am. **149** (3), March 2021

Mahmud *et al.*

Additionally, our findings support theoretical oscillatory models and empirical data (Doelling *et al.*, 2019) that suggest induced activity can predict auditory-perceptual processing better than evoked activity. A disadvantage of this data-driven approach is that it is computationally expensive. Nonetheless, our data suggest that induced neural activity plays a more prominent role in describing the perceptual-cognitive process of speech categorization than evoked modes of brain activity (Doelling *et al.*, 2019). In particular, we demonstrate that among these two prominent functional modes and frequency channels characterizing the EEG, induced γ-frequency oscillations best decode the category structure of speech and the strength of the listeners' behavioral identification. In contrast, the evoked activity provides a reliable though weaker correspondence with behavior in all but the α frequency band. Nevertheless, our study only included vowel stimuli. Additional studies are required to examine if our findings generalize to other speech sounds (e.g., consonants) which elicit stronger/weaker categorical percepts (Pisoni, 1973) or those which are more or less familiar to a listener (e.g., native vs nonnative speech; Bidelman and Lee, 2015).

ACKNOWLEDGMENTS

APPENDIX A

1. SVM optimal parameter values

The optimal parameters of the SVM classifier are given in Table IV in different analysis scenarios.

2. Mean accuracy of SVM fivefold cross-validation results

Using whole-brain *evoked* θ, α, β, and γ frequency responses, speech stimuli (e.g., Tk1/5 vs Tk 3) were correctly distinguished at a mean accuracy of 61%–64%. The LH data revealed that θ-, α-, β-, and γ-bands decoded speech stimuli at mean accuracies between ~57%–63%, whereas decoding from the RH was slightly poorer at 57%–61%. The mean accuracy is ~5% less than the best model using the evoked activity.

Using whole-brain *induced* θ, α, β, and γ frequency response speech stimuli were decodable at mean accuracies of 87%–93%. Still, among all induced frequency bands, the γ-band showed the best speech segregation (93% mean accuracy). The LH oscillations decoded speech categories at 75%–85% mean accuracy, whereas the RH yielded 80%–82% mean accuracy. The maximum deviation of accuracy was ~2% from the best model using induced activity (see Fig. 8).

3. Decoding categorical neural responses using band frequency features and KNN

We used a KNN classifier to corroborate the main SVM findings with a different algorithm. We split the data into training and test sets of 80% and 20%, respectively. During the training phase, we tuned the value of k parameters from one to ten to achieve the maximum accuracy. Classification results of the KNN on the test dataset are reported in Fig. 9. The KNN classifier exhibited similar though slightly inferior results than the SVM classifier exhibited, justifying our choice of the SVM model in the main text.

Using whole-brain *evoked* θ, α, β, and γ frequency responses, speech stimuli (e.g., Tk1/5 vs Tk 3) were correctly distinguished at 64%–68% accuracy. Among all evoked frequency bands, the β-band was optimal to decode the speech categories (67.70% accuracy). The LH data revealed that θ-, α-, β-, and γ-bands decoded speech stimuli at accuracies between ~58%–63%, whereas decoding from the RH was slightly poorer at 57%–62% accuracy.

Using whole-brain *induced* θ, α, β, and γ frequency responses, the speech stimuli were decodable at accuracies of 88%–94%. Among all induced frequency bands, the γ-band showed the best speech segregation (94.42% accuracy). Hemisphere-specific data again showed lower accuracy. The LH oscillations decoded speech categories at 76%–86% accuracy, whereas the RH yielded 79%–84% accuracy.

TABLE IV. Optimal parameters (e.g., C and γ) of the SVM.

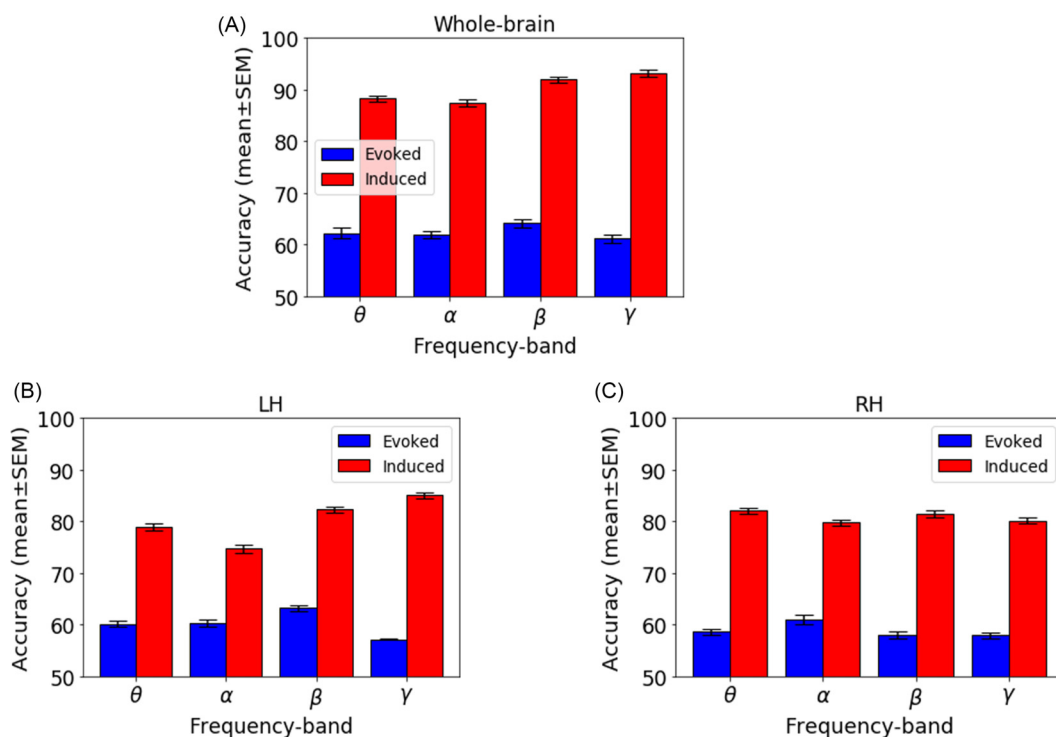| Neural activity	Frequency band	Full brain	LH	RH
Evoked	θ	$C = 30, \gamma = 0.001$	$C = 20, \gamma = 0.001$	$C = 10, \gamma = 0.004$
	α	$C = 40, \gamma = 0.0001$	$C = 20, \gamma = 0.001$	$C = 20, \gamma = 0.003$
	β	$C = 40, \gamma = 0.001$	$C = 30, \gamma = 0.002$	$C = 20, \gamma = 0.002$
	γ	$C = 10, \gamma = 0.0001$	$C = 30, \gamma = 0.003$	$C = 20, \gamma = 0.002$
Induced	θ	$C = 20, \gamma = 0.001$	$C = 10, \gamma = 0.001$	$C = 10, \gamma = 0.001$
	α	$C = 20, \gamma = 0.001$	$C = 20, \gamma = 0.001$	$C = 20, \gamma = 0.001$
	β	$C = 40, \gamma = 0.001$	$C = 40, \gamma = 0.001$	$C = 40, \gamma = 0.001$
	γ	$C = 30, \gamma = 0.001$	$C = 30, \gamma = 0.001$	$C = 30, \gamma = 0.001$

FIG. 8. (Color online) Decoding categorical neural encoding using different frequency band features of the source-level EEG. Mean accuracy of the SVM fivefold cross-validation results classifying prototypical (Tk1/5) vs ambiguous (Tk 3) speech sounds. (A) Whole-brain data (e.g., 68 ROIs), (B) LH (e.g., 34 ROIs), and (C) RH (e.g., 34 ROIs). Chance level = 50%. Errorbars = ±1 s.e.m.
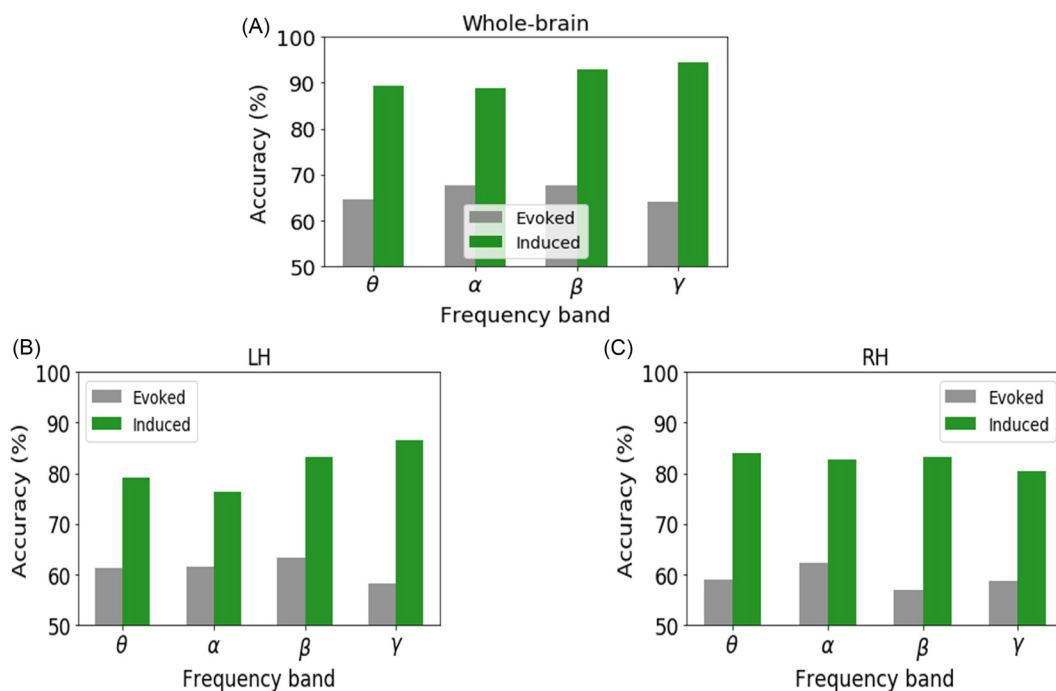

FIG. 9. (Color online) Decoding categorical neural encoding using different frequency band features of the source-level EEG. The KNN results classifying prototypical (Tk1/5) vs ambiguous (Tk 3) speech sounds. (A) Whole-brain data (e.g., 68 ROIs), (B) LH (e.g., 34 ROIs), and (C) RH (e.g., 34 ROIs).

4. Stability selection

Stability selection is a state-of-the-art feature selection method that works well in high dimensional or sparse data based on the Lasso method (Meinshausen and Bühlmann, 2010; Yin et al., 2017). Stability selection can identify the most stable (relevant) features out of a large number of features over a range of model parameters even if the necessary conditions required for the original Lasso method are violated (Meinshausen and Bühlmann, 2010).

In the stability selection, a feature is considered to be more stable if it is more frequently selected over repeated subsamplings of the data (Nogueira *et al.*, 2017). Basically, the randomized Lasso randomly subsamples the training data and fits an L1 penalized logistic regression model to optimize the error. Over many iterations, feature scores are (re)calculated. The features are shrunk to zero by multiplying the features' coefficient by zero while the stability score is lower. Surviving nonzero features are considered important variables for classification. Detailed interpretation and mathematical equations of the stability selection are explained in Meinshausen and Bühlmann (2010). The stability selection solution is less affected by the choice of the initial regularization parameters. Consequently, it is extremely general and widely used in high dimensional data even when the noise level is unknown.

Alain, C. (**2007**). "Breaking the wave: Effects of attention and learning on concurrent sound perception," Hear. Res. **229**, 225–236.

Al-Fahad, R., Yeasin, M., and 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.

Bashivan, P., Bidelman, G. M., and Yeasin, M. (**2014**). "Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity," Eur. J. Neurosci. **40**, 3774–3784.

Bidelman, G. M. (**2015**). "Induced neural beta oscillations predict categorical speech perception abilities," Brain Lang. **141**, 62–69.

Bidelman, G. M. (**2017**). "Amplified induced neural oscillatory activity predicts musicians' benefits in categorical speech perception," Neuroscience **348**, 107–113.

Bidelman, G. M., and Alain, C. (**2015**). "Hierarchical neurocomputations underlying concurrent sound segregation: Connecting periphery to percept," Neuropsychologia **68**, 38–50.

Bidelman, G. M., Bush, L., and Boudreaux, A. (**2020**). "Effects of noise on the behavioral and neural categorization of speech," Front. Neurosci. **14**, 1–13.

Bidelman, G. M., and 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., Mahmud, M. S., Yeasin, M., Shen, D., Arnott, S. R., and Alain, C. (**2019**). "Age-related hearing loss increases full-brain connectivity while reversing directed signaling within the dorsal–ventral pathway for speech," Brain Struct. Funct. **224**, 2661–2676.

Bidelman, G. M., Moreno, S., and Alain, C. (**2013**). "Tracing the emergence of categorical speech perception in the human auditory system," Neuroimage **79**, 201–212.

Bidelman, G. M., and Walker, B. (**2019**). "Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network," NeuroImage **201**, 116022.

Bidelman, G. M., and Walker, B. S. (**2017**). "Attentional modulation and domain-specificity underlying the neural organization of auditory categorical perception," Eur. J. Neurosci. **45**, 690–699.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Anders, M. D., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J. (**2006**). "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," Neuroimage **31**, 968–980.

Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (**2017**). "Cortical alpha oscillations predict speech intelligibility," Front. Human Neurosci. **11**, 1–10.

Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., and Poeppel, D. (**2019**). "An oscillator model better predicts cortical entrainment to music," Proc. Nat. Acad. Sci. U.S.A. **116**, 10113–10121.

Dufor, O., Serniclaes, W., Sprenger-Charolles, L., and Démonet, J.-F. (**2007**). "Top-down processes during auditory phoneme categorization in dyslexia: A PET study," Neuroimage **34**, 1692–1707.

Durgesh, K. S., and Lekha, B. (**2010**). "Data classification using support vector machine," J. Theor. Appl. Inform. Technol. **12**(1), 1–7.

Feng, G., Gan, Z., Wang, S., Wong, P. C., and Chandrasekaran, B. (**2018**). "Task-general and acoustic-invariant neural representation of speech categories in the human brain," Cerebral Cortex **28**, 3241–3254.

Garavan, H., Ross, T. J., and Stein, E. A. (**1999**). "Right hemispheric dominance of inhibitory control: An event-related functional MRI study," Proc. Nat. Acad. Sci. U.S.A. **96**, 8301–8306.

Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., and Laufs, H. (**2007**). "Endogenous cortical rhythms determine cerebral specialization for speech perception and production," Neuron **56**, 1127–1134.

Giraud, A.-L., and Poeppel, D. (**2012**). "Cortical oscillations and speech processing: Emerging computational principles and operations," Nat. Neurosci. **15**, 511–517.

Haenschel, C., Baldeweg, T., Croft, R. J., Whittington, M., and Gruzelier, J. (**2000**). "Gamma and beta frequency oscillations in response to novel auditory stimuli: A comparison of human electroencephalogram (EEG) data with *in vitro* models," Proc. Nat. Acad. Sci. U.S.A. **97**, 7645–7650.

Herrmann, C. S., Rach, S., Vosskuhl, J., and Strüber, D. (**2014**). "Time–frequency analysis of event-related potentials: A brief tutorial," Brain Topogr. **27**, 438–450.

Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., and Giraud, A.-L. (**2015**). "Speech encoding by coupled cortical theta and gamma oscillations," Elife **4**, e06213.

Klimesch, W. (**2012**). "Alpha-band oscillations, attention, and controlled access to stored information," Trends Cognit. Sci. **16**, 606–617.

Luo, H., and Poeppel, D. (**2012**). "Cortical oscillations in auditory perception and speech: Evidence for two temporal windows in human auditory cortex," Front. Psychol. **3**, 170.

Mahmud, M. S., Ahmed, F., Yeasin, M., Alain, C., and Bidelman, G. M. (**2020a**). "Multivariate models for decoding hearing impairment using EEG gamma-band power spectral density," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom, pp. 1–7.

Mahmud, M. S., Yeasin, M., and Bidelman, G. (**2020b**). "Data-driven machine learning models for decoding speech categorization from evoked brain responses," (published online, 2020).

Mankel, K., Barber, J., and Bidelman, G. M. (**2020**). "Auditory categorical processing for speech is modulated by inherent musical listening skills," NeuroReport **31**, 162–166.

Meinshausen, N., and Bühlmann, P. (**2010**). "Stability selection," J. R. Stat. Soc. Series B Stat. Methodol. **72**, 417–473.

Moinuddin, K. A., Yeasin, M., and Bidelman, G. M. (**2019**). *BrainO*, available at https://github.com/cvpia-uofm/BrainO (Last viewed 9/9/2019).

Morillon, B., Liégeois-Chauvel, C., Arnal, L. H., Bénar, C. G., and Giraud, A.-L. (**2012**). "Asymmetric function of theta and gamma activity in syllable processing: An intra-cortical study," Front. Psychol. **3**, 248, 1–9.

Nogueira, S., Sechidis, K., and Brown, G. (**2017**). "On the stability of feature selection algorithms," J. Mach. Learn. Res. **18**(1), 6345–6398.

Ou, J., and Law, S.-P. (**2018**). "Induced gamma oscillations index individual differences in speech sound perception and production," Neuropsychologia **121**, 28–36.

Pisoni, D. B. (**1973**). "Auditory and phonetic memory codes in the discrimination of consonants and vowels," Percept. Psychophys. **13**, 253–260.

Price, C. N., Alain, C., and Bidelman, G. M. (**2019**). "Auditory-frontal channeling in α and β bands is altered by age-related hearing loss and relates to speech perception in noise," Neuroscience **423**, 18–28.

Ruppert, D., and Wand, M. P. (**1994**). "Multivariate locally weighted least squares regression," Ann. Stat. **22**, 1346–1370.

Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitl, M. D., and Liebenthal, E. (**2008**). "Attentional and linguistic interactions in speech perception," Neuroimage **39**, 1444–1456.

Shahin, A. J., Picton, T. W., and Miller, L. M. (**2009**). "Brain oscillations during semantic evaluation of speech," Brain Cognit. **70**, 259–266.

Si, X., Zhou, W., and Hong, B. (**2017**). "Cooperative cortical network for categorical processing of Chinese lexical tone," Proc. Natl. Acad. Sci. U.S.A. **114**, 12303–12308.

J. Acoust. Soc. Am. **149** (3), March 2021

Mahmud *et al.*   1655

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (**2011**). "Brainstorm: A user-friendly application for MEG/EEG analysis," Comput. Intell. Neurosci. **2011**, 1–13.

Tallon-Baudry, C., and Bertrand, O. (**1999**). "Oscillatory gamma activity in humans and its role in object representation," Trends Cognit. Sci. **3**, 151–162.

Yellamsetty, A., and Bidelman, G. M. (**2018**). "Low-and high-frequency cortical brain oscillations reflect dissociable mechanisms of concurrent speech segregation in noise," Hear. Res. **361**, 92–102.

Yin, Q.-Y., Li, J.-L., and Zhang, C.-X. (**2017**). "Ensembling variable selectors by stability selection for the Cox model," Comput. Intell. Neurosci. **2017**, 1–10.