Phonological Contrasts Are Maintained Despite Neutralization: an Intracranial EEG Study

Anna Mai[1], Stephanie Riès[2,3], Sharona Ben-Haim[4], Jerry Shih[5], and Timothy Gentner[6,7,8]

[1]*UCSD, Linguistics,* [2]*SDSU, School of Speech, Language, and Hearing Sciences,* [3]*SDSU, Center for Clinical and Cognitive Neuroscience,* [4]*UCSD, Neurosurgery,* [5]*UCSD, Neurosciences,* [6]*UCSD, Psychology,* [7]*UCSD, Biological Sciences Division, Neurobiology Section,* [8]*UCSD, Kavli Institute for Brain and Mind*

1 Introduction

The existence of language-specific abstract sound-structure units (such as the phoneme) is largely uncontroversial in phonology. However, whether the brain performs abstractions comparable to those assumed in phonology has been difficult to ascertain. Using intracranial electroencephalography (EEG) recorded during a passive listening task, this study takes advantage of several phonological patterns of English to provide evidence that the brain abstracts phonemic and morphemic category identity in a structured, language-specific way from contextually conditioned acoustic variants of phonemes and morphemes (i.e., allophones and allomorphs).

Previous intracranial work on the neural basis of speech sound processing has established that activity in higher auditory areas is sensitive not only to acoustic features of speech, but also to various aspects of the the speech sounds' context. Contexts that have been shown to modulate responses to speech sounds include their degree of intelligibility (Nourski et al., 2019), lexical sequence probability (Leonard et al., 2016), phonological neighborhood density (Cibelli et al., 2015), and the listener's degree of attention to the speech stream (Mesgarani & Chang, 2012). In light of these forms of context-sensitive activity, this study investigates the extent to which abstract phonological and morphological context influence the neural response to speech. In particular, using language-specific (morpho)phonological alternations where surface similarity and abstract underlying identity diverge, we identify neural sites sensitive to either surface similarity or underlying similarity and assess whether the distribution of these sites is unexpected to be observed by chance.

The primary phonological opposition considered is that of /d/ and /t/ and their contextual neutralization to [ɾ]; and the morphophonological processes considered are the formation of the regular past tense and regular plural. These phenomena were chosen to guide the investigation because each involves alternation and neutralization of contrast and occurs with relative frequency in typical English speech. In the case of the the phonological neutralization of coronal stops to tap, the neutralization context provides a window on a one-to-many mapping of acoustic forms to phonological constructs, while the morphophonological alternations provide a window on a slightly more abstract many-to-one mapping from phonological forms to morphological exponence.

English /d/ and /t/ have many acoustically distinct allophones, but when either /d/ or /t/ occurs following a stressed syllable and between two vowels, their acoustic contrast is neutralized, and both are pronounced as a coronal tap (e.g. *writing*, *riding*). Given that auditory processing primarily proceeds in a feedforward manner from spectrotemporal features of the sensory input, it is anticipated that many speech selective sites will demonstrate an acoustic 'surface response', where the responses for all taps are more similar to one another than to the voiceless coronal stop allophone of /t/. However, if it is the case that phonological context is used to compute phonemic identity during language processing, even when the acoustic contrast between two phonemes is neutralized, then there also should exist sites demonstrating a phonemic 'underlying response', where the neural response to underlyingly /t/ taps (**dx_t**; i.e., *writing*) is more similar to other allophones of /t/ (**t**; i.e., voiceless alveolar stops) than to underlyingly /d/ taps (**dx_d**; i.e., *riding*).

Similarly, the two morphophonological contexts of interest also exhibit variation in their realization depending on the phonological context. The regular past tense takes one of three forms: a syllabic voiced

* Thanks to Eric Baković, the members of Gentner Lab, and the AMP 2021 poster session audience for questions, comments, and suggestions on this work. Work was supported by NIMH training fellowship T32MH020002 and William Orr Dingwall Dissertation Fellowship to A.M. Remaining errors are ours.

Proceedings of *AMP 2021*

Patient	Age	Gender	Handed.	Wada	Coverage	Language Experience
SD010	28	M	R	–	sEEG: LH,RH	English
SD011	32	M	R	LH	sEEG: LH,RH	English, Creole, French
SD012	25	F	R	LH	sEEG: LH,RH; Grid,strips: LH	English
SD013	43	M	L	LH*	sEEG: LH,RH	English, Spanish
SD015	55	F	R	–	sEEG: LH,RH	English
SD017	21	M	R	–	sEEG: RH	English, Spanish
SD018	42	F	L	RH	sEEG: LH,RH	English, Spanish
SD019	21	M	R	–	sEEG: LH,RH	English
SD021	33	F	R	–	sEEG: LH,RH	English
SD022	23	M	R	–	sEEG,Grid: RH	English

Table 1: Summary of basic patient information. *While patient SD013 did not experience a clear, prolonged speech arrest with either injection, he demonstrated greater language deficits following left hemisphere injection.

coronal stop [ɘd] following [t] or [d], a voiceless coronal stop [t] following the remaining voiceless consonants, or a voiced coronal stop [d] following the remaining voiced segments. The regular plural analogously manifests in one of three forms: a syllabic voiced coronal sibilant [ɘz] following sibilants [s, z, tʃ, dʒ, ʃ, ʒ], a voiceless coronal sibilant [s] following voiceless non-sibilants, or a voiced coronal sibilant [z] following voiced non-sibilants. Again, since similiarity of acoustic form and neural response is a well-established principle of auditory processing (e.g., Mesgarani et al. 2014), it is anticipated that many speech selective sites will demonstrate a 'surface response', where responses for [z] forms of the plural and [d] forms of the past tense are more similar to word-final non-plural [z] and non-past [d], respectively, than to the voiceless forms of the plural and past tense, respectively. Additionally, if it is the case that morphophonological identity is abstracted from phonological context, then there should also exist sites demonstrating a morphological 'underlying response', where the neural responses to both voiced and voiceless forms of the plural and past tense pattern together to the exclusion of non-plural and non-past word final [z] and [d], respectively.

In this way, observing the distribution of surface and underlying sites for the coronal tap, plural, and past tense comparisons has the potential to provide critical evidence for the mental reality of phonemes and morphemes.

2 Methods

2.1 *Participants* Study participants were ten patients at UC San Diego Health who underwent intracranial stereo electrode implantation as part of treatment for refractory epilepsy. All were native English speakers , and all reported normal hearing and performed within the acceptable range on a battery of neuropsychological language tests. Basic participant information is summarized in Table 1. The research protocol was approved by the UC San Diego Institutional Review Board, and all subjects gave written informed consent prior to surgery.

2.2 *Stimuli* Participants passively listened to short excerpts of conversational American English speech taken from the Buckeye Corpus (Pitt et al., 2007). To assess participant attention to the task, participants responded orally to a two-alternative question about the content of each English passage after they listened to it.

Passages were 25–76s long (mean 38s) taken from 27 (12 women; 15 men) native speakers of Midwestern American English living in Columbus, OH between October 1999 and May 2000. Per corpus documentation (Pitt et al., 2007), excerpts were recorded monophonically on a head-mounted microphone (Crown CM-311A) and fed to a DAT recorder (Tascam DA-30 MKII) at a 48kHz sampling rate through an amplifier (Yamaha MV 802). Following each excerpt, participants heard a two-alternative choice question regarding the content of the passage they heard, and they were asked to respond orally to the question to ensure that they were alert and paying attention to the passages that they heard. These questions were recorded with a Blue Yeti USB microphone sampling at a rate of 48kHz by a speaker of American English. The amplitude of all English passages and questions was normalized to -20dBFS prior to padding with 500ms of silence.

All auditory stimuli were orthographically and phonetically transcribed, segmented, and labeled. Transcription, segmentation, and labeling procedures for passages from the Buckeye Corpus are described in Pitt

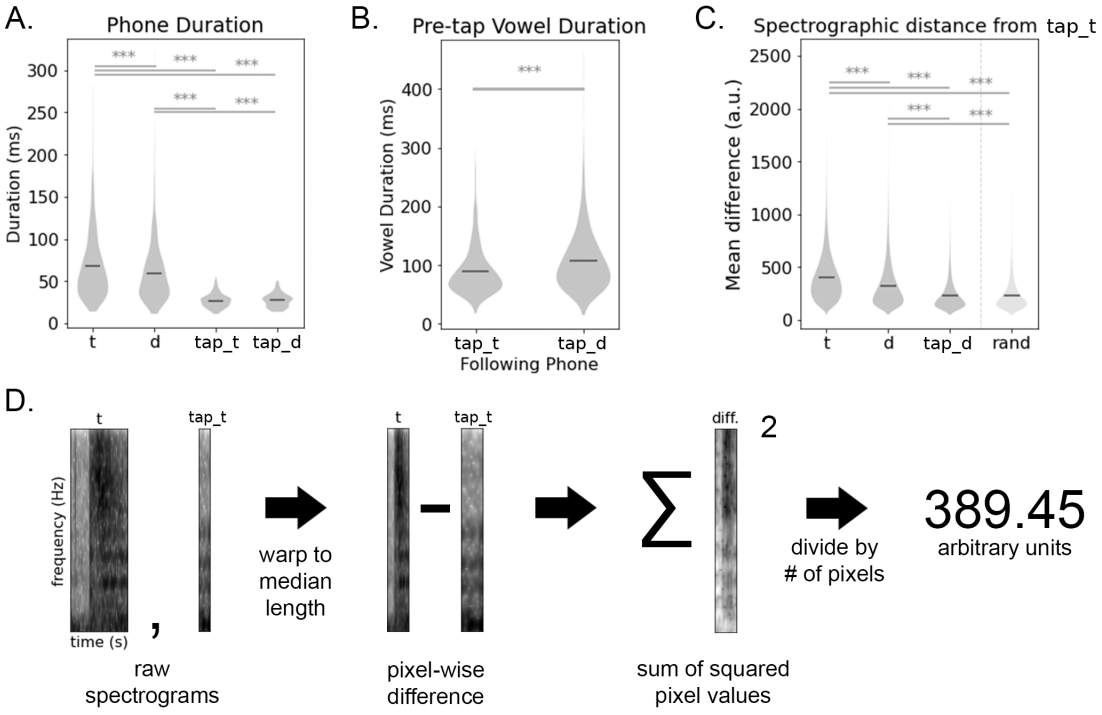

Figure 1: (A) Durations of coronal stops and taps in the Buckeye Corpus. (B) Durations of vowels immediately preceding coronal taps in the Buckeye Corpus. (C) Spectrographic distance of **t**, **d**, and **tap_d** sounds from **tap_t** (left three violins); Spectrographic distance of a random split of all taps from the remainder of taps (rightmost violin). (D) Schematic representation for the calculation of spectrographic distance. Distribution means are shown by the thick dark grey lines in plots A-C, and significance levels are indicated on the following scale: * ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001.

et al. (2007), and transcription, segmentation, and labeling of task instructions and content questions was performed by a phonetically-trained researcher at UC San Diego, using the protocols detailed in Pitt et al. (2007).

2.3 *Coronal Tap Acoustics* The central logic of the study requires coronal taps to be acoustically distinct from coronal stops, and requires there to be no significant acoustic difference between coronal taps based on their phonemic identity. This section demonstrates that this is indeed the case for the coronal stops and taps present in the Buckeye stimuli.

The primary acoustic feature that distinguishes coronal stops from coronal taps is their duration (Zue & Laferriere, 1979; Braver, 2013; Derrick & Schultz, 2013). While voiced and voiceless American English stops are respectively roughly 70–90ms and 98–130ms in duration (Hillenbrand et al., 1984; Hogan & Rozsypal, 1980; Revoile et al., 1982), coronal taps last a mere 10–40ms (Zue & Laferriere, 1979). Previous investigation into the neutralization of /t/ and /d/ in intervocalic environments has found that the variability in coronal tap duration is primarily dependent on the quality of the preceding vowel and not on the phonemic identity of the tap itself (Zue & Laferriere, 1979). In fact, numerous studies have found that /d/ and /t/ taps do not differ significantly in their closure duration (Charles-Luce, 1997; Zue & Laferriere, 1979; Fox & Terbeek, 1977; Sharf, 1962). This finding is replicated in the stimuli used in this study, as shown in Figure 1A. A one-way ANOVA indicated that there was a significant effect of phonemic category on phone duration [$F(3;\ 1{,}666) = 107.75$, $p \leq 0.001$], and *post hoc* comparisons using the Tukey HSD test indicated that voiceless coronal stops in the stimuli are significantly longer than voiced stops, which are in turn significantly longer than coronal taps, each at the $p \leq 0.001$ level. However, /d/ and /t/ taps do not differ significantly from one another in duration ($p = 0.9$). Thus, on the basis of their duration, coronal taps are distinct from coronal stops yet not distinct from one another on the basis of their phonemic identity, satisfying the study's foundational assumptions.

Nevertheless, the stimuli used in this study were taken from natural speech and undoubtedly vary along more acoustic dimensions than duration. As a means of further ensuring that the acoustic differences between coronal stops is greater than the difference between coronal taps, a measure of spectrographic distance was calculated for coronal stops and taps. First, wide-band spectrograms were created for all stimuli using Praat's default settings, and the spectrographic segments corresponding to coronal stops and taps were excised using the segmentation and labelling given in the Buckeye Corpus. The median duration of these segments was calculated (in bins) and all spectrographic segments were resized to this duration using the resize function from the transform module of the Python library scikit-image (van der Walt et al., 2014). This function performs bi-linear interpolation to resize the image. Prior to down-scaling any images, it applies a Gaussian filter with a kernel size of $(s-1)/2$, where s is the down-scaling factor, to prevent any anti-aliasing artifacts. For each pair of resized spectrograms, the sum of the squared pixel-wise difference was calculated, and divided by the size of the spectrogram in pixels. This process is shown schematically in Figure 1D. Figure 1C plots the distances of **t**, **d**, and **tap_d** spectrograms from **tap_t** spectrograms (darker grey), as well as the distances for a random split of all coronal taps (lighter grey). A one-way ANOVA indicated that there was a significant effect of phonemic category on spectrographic distance from **tap_t** [$F(3; 339{,}842) = 9124.47$, $p \leq 0.001$], and *post hoc* comparisons using the Tukey HSD test indicated that the distance between **tap_t** and **tap_d** is significantly less than the distance between **tap_t** and either of the coronal stops at the $p \leq 0.001$ level. However, when mean distance is calculated for a random split of taps not based on phonemic identity, it is not significantly different from the mean distance calculated based on the phonemic identity of the taps ($p = 0.44$). Once again, this shows that the phonemic identity of coronal taps cannot be determined from the acoustic properties of taps themselves.

Although taps derived from /d/ and /t/ are acoustically indistinguishable from one another, the length of the vowel that precedes a tap systematically varies in duration based on the phonemic identity of the tap. When preceding a tap derived from /d/, vowels are approximately 10% longer than those preceding a tap derived from /t/ (Sharf, 1962; Zue & Laferriere, 1979; Herd et al., 2010; Braver, 2013). This pattern is also observed in the Buckeye Corpus, where vowels preceding taps derived from /d/ were on average 18.31ms (SD=[13.23, 23.38]) longer than vowels preceding taps-derived from /t/ [$t(1309) = 50.16$, $p \leq 0.001$] (Figure 1B). Thus, there does exist an acoustic cue to the phonemic identity of taps. On this basis, one could argue that any observed difference in the neural response to medial /d/ and /t/ is due to the duration of the preceding vowel, seemingly undermining the foundational assumption of the study. For this reason, this study focuses on the 500ms following the onset of coronal stops and taps. By timelocking the analyzed response to the beginning of closure and using the preceding 100ms to baseline the signal, the acoustic impact of the preceding vowel is effectively neutralized. Though it remains possible that the duration of the preceding vowel cues the listener to the phonemic identity of the following tap, differences observed in the response /d/ and /t/ taps must be based on their categorization, and not the acoustics of the preceding vowel itself, because the mean signal in the 100ms preceding the tap is subtracted from the analyzed response, and /d/ and /t/ taps themselves do not differ acoustically from one another. In other words, though the duration of the vowel may cue the phonemic identity of the tap, by timelocking the response to the beginning of the tap, any difference in response cannot be reducible to the preceding *acoustic* context.

2.4 *Data Acquisition & Processing* Experimental instructions and stimuli were presented to participants in their hospital rooms on a Windows 10 desktop PC (Dell XPS 8910) using PsychoPy for Python 2.7 (Peirce, 2007, 2008). The task was conducted in six blocks, each of which contained eight English trials. The average block duration was 6 minutes 30 seconds. All participants completed at least two of the six blocks, and the average participant completed four blocks, listening to 26:45 minutes of speech totaling 15,070 phones. Each English trial consisted of a short, auditorily-presented passage followed by a content question and an oral response.

Intracranial EEG signals were amplified using a multi-channel amplifier system (Natus Quantum) and recorded using Natus NeuroWorks software. Auditory stimuli and oral responses were recorded simultaneously with the EEG data by feeding the output of a Zoom H2n microphone as an additional input channel to the Natus Quantum amplifier.

After recording, neural data were deidentified and exported from the clinical NeuroWorks system in .edf (European Data Format) format for pre-processing using the Python package MNE Python (Gramfort et al., 2013). Channels displaying excessive artifacts or line noise were removed (45 channels total). Remaining channels were common average referenced, notch filtered at 60Hz and its harmonics, bandpass filtered 0.1–170Hz, and downsampled to three times the lowpass cutoff (510Hz). Independent Component Analysis (ICA) was used to remove stationary artifacts from the filtered data, and time intervals containing remaining artifacts

were visually identified and discarded. Channels selected for analysis were those which exhibited reliable evoked response to speech stimuli, determined by a sliding window t-test between responses to randomly-selected time frames during the passive listening task and in silence ($p < 0.05$).

2.5 *Band power* The frequency bands used in this study were delta (1–4Hz), theta (4–7Hz), alpha (8–12Hz), beta (13–30Hz), gamma (31–50Hz), and high gamma (70–150Hz) bands. To compute the power for each band, the analytic amplitude from eight Gaussian band-pass filters with logarithmically increasing center frequency (Crone et al., 1998; Bouchard et al., 2013) was averaged. For analysis, each band-passed signal was then segmented into peri-target epochs with 100ms pre-target and 500ms post-target. Each epoch was then z-scored relative to the mean and standard deviation of its 100ms pre-target baseline, where the target is the onset of a phone of interest.

3 Results

3.1 *Significant Electrodes* Broadband neural activity was recorded from a total of 1,355 valid electrodes[1] across the ten subjects, and each electrode was assessed for speech selectivity using a sliding-window one-way t-test, where 500ms silent epochs were compared against an equivalent number of randomly sampled speech epochs. Within each group, epochs were z-scored against a 100ms baseline and t-tests were calculated for nine 100ms windows with 50ms overlap across the 500ms non-baseline portion of the epoch. Within each window, values for each token were averaged over time. Channels were considered speech selective if the t-test for at least one window was significant with $p < 0.05$.

Speech-selective electrodes were defined independently for each band, and across the ten subjects, 1,091 electrodes were found to be speech selective for at least one band, with an average of 339 electrodes found to be speech selective for each band. While some overlap was observed in the sites categorized as speech-selective across bands, the majority of speech-selective sites were speech-selective for only one band. Phonological and morphophonological comparisons were then performed only for speech selective electrodes.

3.1.1 *Phonological Comparisons* Acoustic surface sites and phonemic underlying sites were defined using a sliding-window one-way ANOVA. For a site to be considered an acoustic surface site, there must have existed at least one time window with a significant ANOVA for which a Tukey's *post hoc* test indicated that there was a significant difference ($\alpha = 0.05$) between **t** and **tap_t** tokens and between **t** and **tap_d** tokens but no significant difference between **tap_t** and **tap_d** tokens. Alternatively, for a site to be considered a phonemic site, there must have existed at least one time window with a significant ANOVA for which a Tukey's *post hoc* test indicates that there was a significant difference between **tap_d** and **tap_t** tokens and between **tap_d** and **t** tokens but no significant difference between **t** and **tap_t** tokens.

Of the roughly 339 speech selective electrodes for each band, an average of 5 acoustic sites (SD$\pm$4.7) and 30 phonemic sites (SD$\pm$6.3) were observed for the coronal stop–tap alternation across all bands. The number of acoustic and phonemic sites observed across bands is summarized in Table 2, and examples of the response observed at acoustic and phonemic sites are shown in Figure 2. Only one site was categorized as both an acoustic and phonemic site. This site, RA11, was recorded in white matter beneath right superior temporal sulcus in patient SD022. The phonemic response at this site occurred at 50–150ms after target onset, preceding the acoustic response at 350–450ms, and the effect was observed only in delta band power.

To assess the likelihood of observing these numbers of acoustic and phonemic response sites by chance, an expected null distribution was generated for each frequency band by performing the statistical analysis described above using 1,000 arbitrary pairs of phones (i.e., **A**, **B**) with an arbitrary split of one phone (i.e., **A**, **B_x**, **B_y**). For each each arbitrary set of phones, pseudo surface sites were identified as those with at least one time window in which there was a significant difference between the evoked response to **A** phones and the evoked response to **B** phones, but no significant difference in the evoked response to **B_x** and **B_y** phones. Pseudo underlying sites were identified as those with at least one time window with no significant difference between **A** and **B_x** phones but a significant difference between those phones and **B_y** phones. In this way, the null distribution was generated from the real, recorded data. Spatially correlated activity is thus preserved in the null distribution, accounting for the possibility that such correlated activity could inflate the number of observed surface and underlying sites.

[1] That is, electrodes not discarded due to the presence of excessive line noise or artifacts.

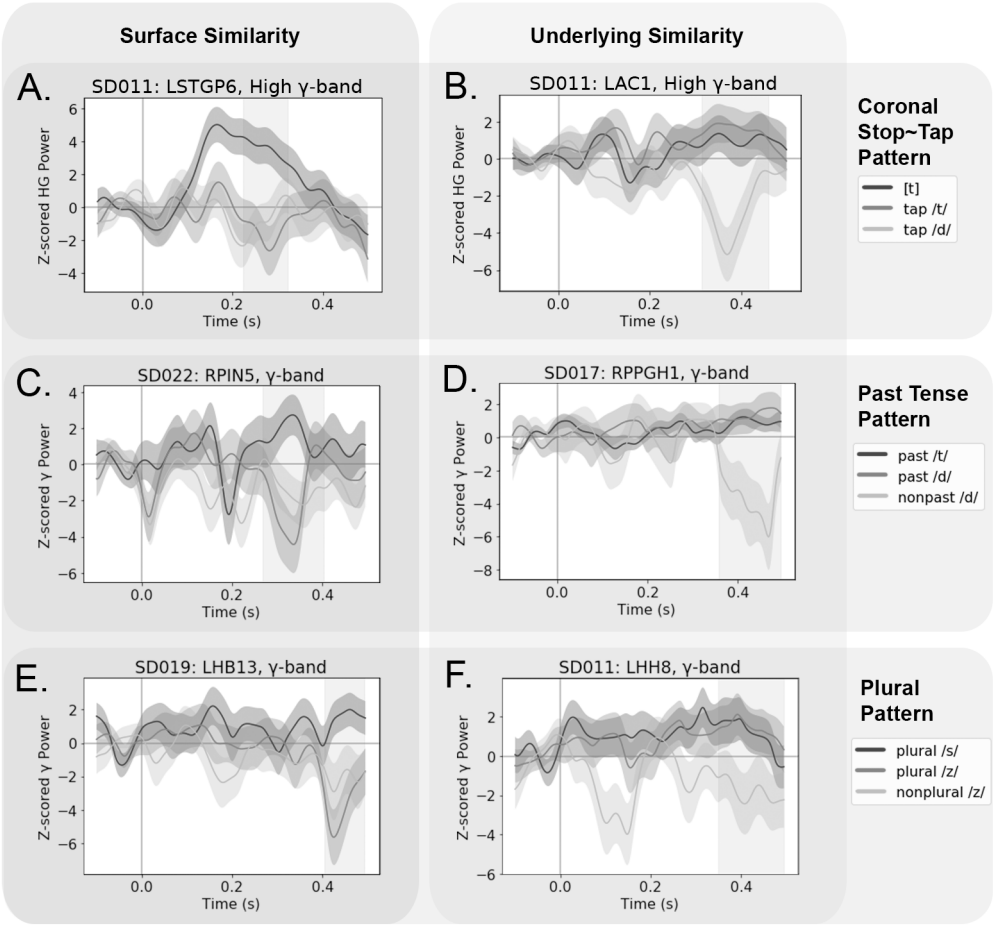

Figure 2: Responses to coronal tap, plural, and past tense comparisons exhibit both surface similarity and underlying similarity patterns. Subplot titles indicate the subject identity, channel name, and response band being plotted, and each shows the time course of band power z-scored relative to baseline (-100ms–0ms). (A) and (B) show sites identified by the coronal stop-tap alternation. The evoked response to tokens of [t] is shown in the darkest grey; the evoked response to taps derived from /t/ (tap /t/) is shown in the middle grey; and the evoked response to taps derived from /d/ (tap /d/) is shown in the lightest grey. (C) and (D) show sites identified by the past tense alternation. The evoked response to /t/ allomorphs of the past tense is shown in the darkest grey; the evoked response to /d/ allomorphs of the past tense is shown in the middle grey; and the evoked response to word-final nonpast tokens of /d/ is shown in the lightest grey. (E) and (F) show sites identified by the plural alternation. The evoked response to /s/ allomorphs of the plural is shown in the darkest grey; the evoked response to /z/ allomorphs of the plural is shown in the middle grey; and the evoked response to word-final nonplural tokens of /z/ is shown in the lightest grey. For all subplots, shading indicates ±SEM.

Compared against these null distributions, the numbers of observed sites exhibiting surface and underlying patterns of activity were greater than would be expected by chance for each band. For example, for the high gamma band, based on the generated distribution, the probability of observing at least 23 phonemic underlying sites and at least 11 acoustic surface sites was $<1\%$. Figure 3 shows the null distributions of surface and underlying sites, with a dotted white line marking the bound for 95% of the distribution's mass and circular points indicating the observed number of surface and underlying sites.

3.1.2 *Morphophonological Comparisons* For the two morphophonological comparisons, sites were also classified as surface or underlying using a sliding-window one-way ANOVA assessment similar to that described for the purely phonological coronal stop alternation. For the regular past tense, surface sites were

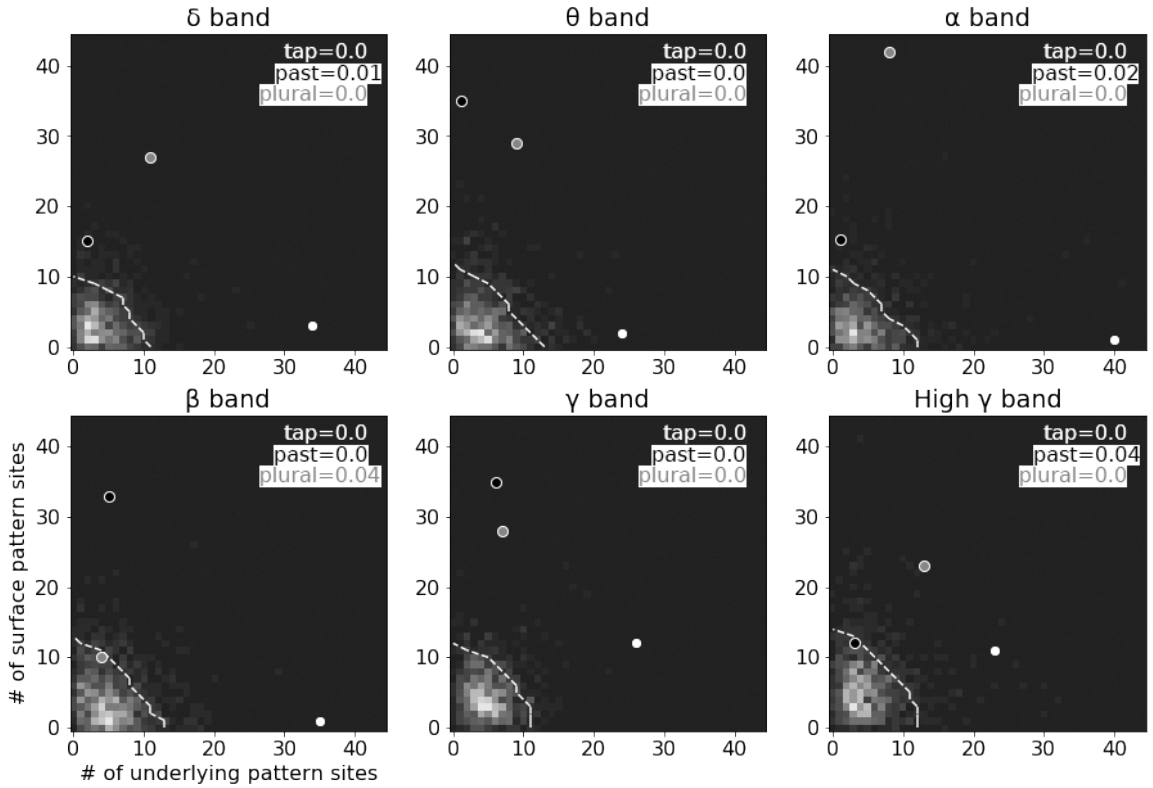

Figure 3: Surface similarity and underlying similarity patterns are not random. Number of significant sites observed for each (morpho)phonological comparison for each neural response band relative to the generated null distribution for that band. Vertical axes indicate the number of sites selective for surface identity observed for each comparison, while horizontal axes indicate the number of sites selective for underlying identity observed for each comparison. The proportion of the null distribution that contains at least as many surface and underlying sites as were observed for each comparison is indicated in the top right corner of each plot. Values for the tap comparison are white; values for the past tense comparison are black encircled in white; and values for the plural comparison are grey encircled in white. Dashed white lines delimit the boundary containing 95% of the null distribution.

considered to be those where the sliding-window ANOVA was significant for at least one time window and for which a Tukey's *post hoc* test indicated that there was a significant difference ($\alpha = 0.05$) between past tense **t** and past tense **d** tokens and between past tense **t** and word final non-past **d** tokens but no significant difference between past tense **d** and word final non-past **d** tokens. Similarly, for the regular plural, surface sites were defined as those with at least one time window for which a Tukey's *post hoc* test indicated that there was a significant difference ($\alpha = 0.05$) between plural **s** and plural **z** tokens and between plural **s** and word final non-plural **z** tokens but no significant difference between plural **z** and word final non-plural **z** tokens.

For the regular past tense alternation, morphological underlying sites were considered to be those for which the comparison of evoked responses to past tense **t**, past tense **d**, and word-final non-past **d** resulted in at least one time window indicating a significant difference between word-final non-past **d** and past tense **t** tokens and between word-final non-past **d** and past tense **d** tokens but no significant difference between past tense **t** and past tense **d** tokens. Similarly, for the regular plural alternation, morphological underlying sites were those for which there was at least one time window indicating a significant difference between word-final non-plural **z** and plural **s** tokens and between word-final non-plural **z** and plural **z** tokens but no significant difference between plural **z** and plural **s** tokens. In this way, morphological underlying sites were those which maintained language-specific morphological similarity based on the meaning of similar word-final sounds rather than maintaining the comparably less abstract similarity of their acoustic or phonemic realization.

For the past tense alternation, across all bands an average of 24.2 (SD$\pm$10.2) were categorized as surface

	Coronal Tap Comparison		*Plural Comparison*		*Past Tense Comparison*	
Band	Surface	Underlying	Surface	Underlying	Surface	Underlying
δ: (1–4Hz)	3	34	27	11	15	2
θ: (4–7Hz)	2	24	29	9	35	1
α: (8–12Hz)	1	40	42	8	15	1
β: (13–30Hz)	1	35	10	4	33	5
γ: (31–50Hz)	12	26	28	7	35	6
High-γ: (70–150Hz)	11	23	23	13	12	3

Table 2: Count of sites demonstrating a surface similarity pattern or an underlying similarity pattern for each frequency band. The observation of more surface sites than underlying sites for the morphological comparison is likely due to the fact that surface similarity for the morphological comparisons is analogous to both surface and underlying similarity at the phonological level.

sites, 3 (SD$\pm$1.9) were categorized as morphological sites, and four sites were categorized as both surface and morphological sites. For the plural alternation, 26.5 (SD$\pm$9.4) sites were categorized as surface sites, 8.7 (SD$\pm$2.9) were categorized as morphological underlying sites, and six sites were categorized as both surface and morphological sites. As for the purely phonological, coronal tap comparison, the numbers of sites exhibiting surface and underlying patterns of activity were greater than would be expected by chance for both the regular past tense and plural comparisons. These likelihoods were assessed against the same null distribution as was used to assess the significance of the phonological coronal stop alternation, a distribution generated by performing this statistical analysis for 1,000 arbitrary pairs of phones (i.e., **A**, **B**) with an arbitrary split of one phone (i.e., **A**, **B_x**, **B_y**).

From this distribution, the probability of observing the past tense pattern of at least 12 surface sites and at least 3 underlying sites was 1.7% for the high gamma band. The probability of observing the plural pattern of at least 23 surface sites and at least 13 underlying sites was less than 1% for the same band. Figure 3 shows the null distribution of surface and underlying sites for each frequency band, with a dotted white line marking the bound for 95% of the distribution's mass and single black (past) and grey (plural) points circled in white indicating the observed number of surface and underlying sites for the two morphophonological comparisons.

4 Discussion

4.1 *Implications for Phonology* The central finding of this study is that more sites sensitive to phonemic identity were observed than would be predicted by chance. While simple in its articulation, this result has far-reaching consequences for our conceptions of phonology and language representation in the brain. In particular, these findings support a reallocation of probability mass away from a number of common ideas about the nature of phonology and phonological processing.

First, implicit in many theories of speech production and processing is the assumption that language-specific grammar only occurs at the level of syntax. Phonological knowledge, including knowledge of language-specific phonotactics, is cast as epiphenomenal of lexical knowledge: speech perception is more or less a word recognition problem, and speech production proceeds in a universal and deterministic manner from whatever articulo-acoustic content is specified in the lexicon. This study provides evidence that these common assumptions are incomplete. Language-specific phonological grammar guides the neural response to speech, and it does so at a level of granularity commensurate with the phoneme.

To see how this conclusion follows from the observed data, consider what one would predict if phonemes were not a relevant unit of organization for language in the brain. To maintain true skepticism about the cognitive reality of phonemes, one would need to envision a state of affairs in which underlying representations of speech sounds were altogether unnecessary. That is, at no point in speech processing would a listener assign the value /d/ to some taps and /t/ to others. Perhaps in lieu of underlying representations, the forms of words that the brain had access to would be exemplars, such as pronunciations the individual had heard before. Laying aside the fact that such a system would likely fail to explain how speakers produce forms they haven't heard before, if it were the case that words containing taps were specified as such in their lexical entries, it would be unexpected to observe any of the sites in this paper that have been called 'phonemic'. Instead, one would expect to see only

acoustic sites.

Having observed phonemic sites, however, if one were to maintain that lexical representations reflect sounds as they are produced, the distinction between /d/-taps and /t/-taps observed in phonemic sites would need to be accounted for by some other means. The most readily available explanation would be orthographic influence. In most all cases, the orthographic forms of words containing taps reliably differentiates words containing taps that are hypothesized to derive from /d/ from those that are hypothesized to derive from /t/. On this basis, one could argue that phonemic sites are indicative of the pervasiveness of orthographic influence on literate language users. There is some evidence to suggest that this level of orthographic influence is possible. For example, using the French minimal triplet *prix* pʁi, *tri* tʁi, *cri* kʁi, during a passive listening task, Pattamadilok et al. (2014) show that a larger Mismatch Negativity (MMN) is elicited when the spelling of the deviant is incongruent with the spelling of the standard (i.e., *prix* vs. *cri*) than when the spelling of the deviant is congruent with the standard (i.e., *cri* vs. *tri*). However, as the authors themselves note, the task of passively listening to strings of repeated words is not particularly natural, and due to the inherent constraints of the MMN and the French lexicon, only these three words were used in the study.

However, in the context of the current study, the likelihood that orthographic influence alone drives the appearance of phonemic underlying sites is low. If it were the case that orthographic influence drove the occurrence of phonemic sites for the tap comparison, we would not expect to see surface sites for the past tense comparison. The regular past tense is consistently spelled with an '-ed' and yet, to observe surface sites for the past tense comparison in this study, pasts produced as [d] pattern with other word final [d] sounds, to the exclusion of pasts produced as [t]. If orthography was primarily responsible for the results of this study, we would not expect to see this structured split between past tense [d] and past tense [t], since orthography is consistent across all forms. A very similar argument can be made from the observation of surface sites for the regular plural comparison.

Perhaps then one would argue that orthography influences perception only when there is very little acoustic difference between two sounds. The [d] and [t] forms of the past tense are sufficiently acoustically distinct from one another, while /t/-taps and /d/-taps are practically indistinguishable. If orthographic knowledge is only called upon to bootstrap lexical recognition when acoustic discriminability is poor, then one could argue that orthographic influence is responsible for the the so-called phonemic sites in the case of the tap comparison, but not for the surface sites of the regular plural and past tense comparisons. This very well may be the case, and is worthy of its own empirical investigation. However, any study which would attempt to show that orthographic influence is responsible for the appearance of phonemic sites in this study must also take care to control for the substantial correlations between orthography and underlying phonemics, for orthographic systems are not randomly structured and often reflect codification of phonological knowledge (see Chomsky 1970 for a discussion of English).

However, if one accepts that orthographic effects alone cannot account for the results observed, the question remains: How can it be that more phonemic sites are observed than would be expected by chance? Here, it is argued that the most parsimonious explanation for this result is language-specific phonological knowledge of the kind generally assumed by working phonologists. More precisely, the existence of phonemic sites suggests that the representations of speech sounds and words are not merely amalgamations of pronunciations that a listener has encountered before. Rather, there is a degree of abstraction between surface, acoustic forms and the prelexical forms that speech sounds are mapped onto in the brain. The nature of this abstraction is language-specific, reflecting not only differences in the phonemic inventories of different languages, but also the language-specific, contextually-sensitive sound alternations that comprise a specific language's phonological grammar.

Furthermore, phonological theories that assume the sole linguistic content of phonemes takes the form of distinctive features are also likely inadequate. If we assume, uncontroversially, that the difference between /t/ and /d/ and the difference between /s/ and /z/ is their value for the feature [±voice], then we would predict that surface sites for the past tense comparison, where tokens of /t/ are distinguished from /d/ regardless of their morphological content, would be the same as surface sites for the plural comparison, where tokens of /s/ and /z/ are distinguished. However, across all subjects and bands, less than 10% of surface sites are shared between the past tense and plural comparisons, as shown in Figure 4. Moreover, when phonemic sites identified from the tap comparison, in which tokens of /d/ and /t/ are also hypothetically distinguished by the single feature [±voice], are considered alongside plural and past tense surface sites, the number of sites shared in common drops to one. Thus, it is highly unlikely that these sites index the presence or absence of the feature [±voice], since we would not expect the addition of phonemic sites identified by the tap comparison to further exclude candidate [±voice] sites that were identified by the past tense and plural surface comparisons. In this way, the

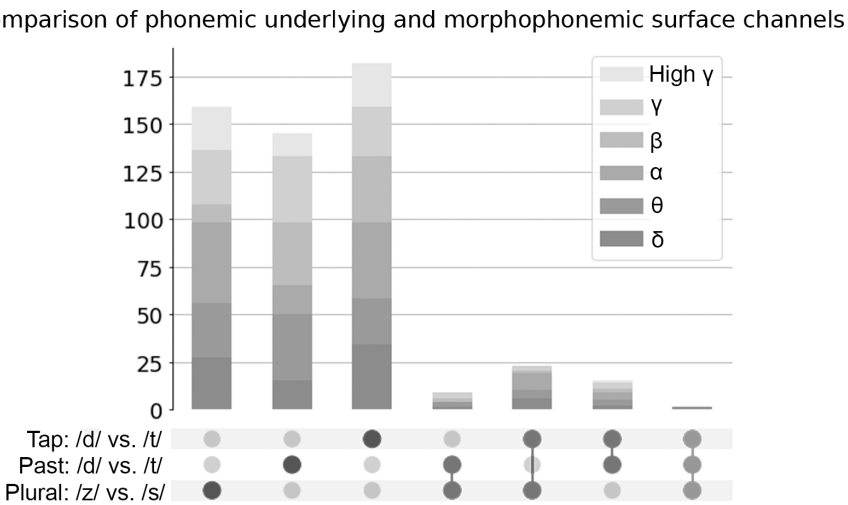

Figure 4: A comparison of which phonemic underlying sites (from the tap comparison) and which morphophonemic surface sites (from the plural and past comparisons) are shared in common across response bands. The lower dot matrix plot indicates the combination of comparisons being considered, and the upper bar plot shows the number of significant channels shared in common for that combination of bands. Colors within each bar indicate the distribution of significant channels across response bands.

findings of this study compel careful reconsideration of the status of distinctive features in phonological theory.

For instance, Dresher (2011) lays out three ways that phonologists have conceptualized of the phoneme as an object of inquiry. One class of definitions characterizes the phoneme as a *physical reality*. Representative definitions of this genre describe the phoneme as a language-specific "family" of sounds that "count for practical purposes as if they were one and the same" (Jones 1967: 258, via Dresher 2011), perhaps because "the speaker has been trained to make sound-producing movements in such a way that the phoneme features will be present in the sound waves, and [the speaker] has been trained to respond only to these features" (Bloomfield 1933: 77-78). Psychological theories of language, such as Forster's (1979), are more or less consistent with this category of explanation, given their focus on direct mappings from physical sounds to lexical entries. The second class of definitions views the phoneme as a *psychological concept*. This class of explanation is often presented as an alternative to the physical reality of the phoneme: if a physical constant coextensive with the phoneme does not exist, then perhaps a mental constant does instead. Finally, if both the physical and psychological reality of the phoneme are rejected, then Dresher (2011), echoing Twaddell (1935), concludes that the phoneme is a convenient *theoretical fiction*, without material basis in the mind, mouth, or middle ear.

In some sense, all three of these positions must be true, in different ways. There must exist some relationship between the physical properties of speech sounds and the roles they play in language, since the experience of spoken language comprehension has a principled relationship to physical speech. Additionally, there must exist some physical substrate in the brain to bridge the sensory periphery with the experience of language comprehension. In the abstract, this bridge may be considered to be composed of mental representations. However, ultimately, not all phonological representations that have been theorized will have equal explanatory adequacy in the brain. Thus, some aspects of the phoneme will remain useful theoretical fictions.

The study presented in this paper addresses the second of these three positions, investigating whether the commonly held understanding of the phoneme as a unit of language-specific contrast has psychological reality or merely theoretical utility. In demonstrating the existence of sites sensitive to phonemic contrast in the absence of acoustic distinction, this study provides support for the classical understanding of the phoneme as a psychological entity and a core level of sublexical linguistic processing. At the same time, it suggests that some theoretical implementations of the phoneme as a unit characterized by minimal (as opposed to exhaustively-specified) sets of distinctive features may have more theoretical utility than psychological reality. To the extent that linguistic and psychological theories of phonology intend to account for the same sets of behavioral phenomena, these results nevertheless support the continued use of phonemic units in phonological theory and

reify the existence of such language-specific sublexical structure in language processing.

4.2 *Implications for Morphology* The morphological results of this study also engage with foundational principles of sublexical language processing. In observing more surface and morphological underlying sites than would otherwise be expected by chance for both the plural and past tense comparisons, this study provides evidence for two interrelated processes in the neural basis of morphology.

It is generally accepted that morphological decomposition takes place during the processing of regularly inflected forms in English and related languages (Münte et al. 1999; Marslen-Wilson & Tyler 2007; Bozic & Marslen-Wilson 2010; Schiller 2020; though cf. Sereno & Jongman 1997). The presence of morphological underlying sites in this study is generally consistent with these results. However, the paradigms typically employed to assess the compositionality of morphological units within words arguably gauge fairly abstract proxies for morphological structure. Many rely on priming, and assess compositionality through metrics such as reaction time or expectation violation response. The evidence for morphological abstraction presented here has the advantage being gathered during a naturalistic listening task and straightforwardly comparing the difference in neural response to sounds that bear morphological exponence to those that do not carry any morphological meaning.

From this perspective, the difference in response between non-plural word-final [z] and plural [s] or [z] can be attributed straightforwardly to the morphological exponence of [s] or [z], without appeal to intermediary cognitive phenomena such as priming or expectation. That is, the meaning of 'plural' associated with the sounds [s] and [z] drives the response of the morphological underlying sites for the plural comparison, and likewise, the meaning of 'past' associated with [t] and [d] drives the response of the morphological underlying sites for the past tense comparison. This, in general, is a great strength of the paradigm employed in this study. However, these results themselves do not provide explicit evidence that the structure of regular plural and past tense forms is compositional, since it could be the case that, for instance, plural [s] and [z] pattern together at morphological sites through an analogical process. In such a case, plural [s] and plural [z] would still evoke similar neural responses because they index a common morphological exponent, but that commonality would be mediated through an analogical process via the lexicon rather than a compositional, syntax-like process within the word itself. Nevertheless, the results of this study provide tantalizing early evidence of the neural response to the presence of particular morphological exponents.

Moreover, these results support the idea that morphological identity is abstracted over phonologically distinct alternants in a structured, language-specific manner. As was argued for the phonemic sites identified by the tap comparison, the structured relationships between speech sounds and their phonological contexts form a language-specific grammar of sounds. Given their interface with meaning, morphophonological alternations in particular have been of central importance to the development of phonological theory (see Kenstowicz & Kisseberth 2014) and are often appealed to in classic texts as foundational evidence for the existence of the phoneme, since it is through the ways that sounds either change or fail to change the meanings of words that phonemic contrast is most strikingly established. In this way, the existence of morphological underlying sites identified by the plural and past tense comparisons demonstrates both that morphological identity is abstracted over distinct, phonologically conditioned alternants and that morphological exponence transcends phonemic particularities.

5 Conclusion

Using a novel experimental design based on the foundational concepts of phonological contrast and neutralization, this study investigates the representation of phonological units in the brain and the relationship between those units, auditory sensory input, and higher levels of language organization, namely morphology. Leveraging the phonological neutralization of coronal stops to tap in English, this study provides evidence of a dissociation between acoustic similarity and phonemic identity across sites recorded intracranially while participants listened to conversational speech. Moreover, leveraging morphophonological alternations of the regular plural and past tense, this study further demonstrates early (< 500ms) evidence of dissociation between phonological form and morphological exponence. Together these results highlight the central nature of language-specific knowledge in sublexical language processing and improve our understanding of the ways language-specific knowledge structures and organizes speech perception in the brain.

References

Bouchard, Kristofer E., Nima Mesgarani, Keith Johnson & Edward F. Chang (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:7441, 327–332.

Bozic, Mirjana & William Marslen-Wilson (2010). Neurocognitive contexts for morphological complexity: Dissociating inflection and derivation. *Lang. Linguist. Compass* 4:11, 1063–1073.

Braver, Aaron (2013). Incomplete neutralization in American English flapping: A production study. *Proc. Mtgs. Acoust.* 19:1.

Charles-Luce, Jan (1997). Cognitive factors involved in preserving a phonemic contrast. *Lang. Speech* 40 (Pt 3), 229–248.

Cibelli, Emily S., Matthew K. Leonard, Keith Johnson & Edward F. Chang (2015). The influence of lexical statistics on temporal lobe cortical dynamics during spoken word listening. *Brain and Language* 147, 66–75.

Crone, Nathan E., Diana L. Miglioretti, Barry Gordon & Ronald P. Lesser (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. event-related synchronization in the gamma band. *Brain* 121 (Pt 12), 2301–2315.

Derrick, Donald & Benjamin Schultz (2013). Acoustic correlates of flaps in North American English. *Proceedings of Meetings on Acoustics ICA2013*, Acoustical Society of America, vol. 19.

Dresher, B. Elan (2011). The phoneme. Van Oostendorp, Marc, Colin J. Ewan, Elizabeth Hume & Keren Rice (eds.), *The Blackwell Companion to Phonology*, John Wiley & Sons, Ltd Oxford, UK, chap. 11, 1–26.

Fox, Robert A. & Dale Terbeek (1977). Dental flaps, vowel duration and rule ordering in american english. *J. Phon.* 5:1, 27–34.

Gramfort, Alexandre, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen & Matti Hämäläinen (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7, p. 267.

Herd, Wendy, Allard Jongman & Joan Sereno (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *J. Phon.* 38:4, 504–516.

Hillenbrand, James, Dennis R. Ingrisano, Bruce L. Smith & James E. Flege (1984). Perception of the voiced–voiceless contrast in syllable-final stops. *J. Acoust. Soc. Am.* 76:1, 18–26.

Hogan, John T. & Anton J. Rozsypal (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *J. Acoust. Soc. Am.* 67:5, 1764–1771.

Kenstowicz, Michael & Charles Kisseberth (2014). *Generative Phonology: Description and Theory*. Academic Press.

Leonard, Matthew K., Maxime O. Baud, Matthias J. Sjerps & Edward F. Chang (2016). Perceptual restoration of masked speech in human cortex. *Nature communications* 7:1, 1–9.

Marslen-Wilson, William D. & Lorraine K. Tyler (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362:1481, 823–836.

Mesgarani, Nima & Edward F Chang (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485:7397, 233–236.

Mesgarani, Nima, Connie Cheung, Keith Johnson & Edward F. Chang (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343:6174, 1006–1010.

Münte, Thomas F., Tessa Say, Harald Clahsen, Kolja Schiltz & Marta Kutas (1999). Decomposition of morphologically complex words in English: evidence from event-related brain potentials. *Brain Res. Cogn. Brain Res.* 7:3, 241–253.

Nourski, Kirill V., Mitchell Steinschneider, Ariane E. Rhone, Christopher K. Kovach, Hiroto Kawasaki & Matthew A. Howard III (2019). Differential responses to spectrally degraded speech within human auditory cortex: an intracranial electrophysiology study. *Hearing research* 371, 53–65.

Pattamadilok, Chotiga, José Morais, Cécile Colin & Régine Kolinsky (2014). Unattentive speech processing is influenced by orthographic knowledge: Evidence from mismatch negativity. *Brain Lang.* 137, 103–111.

Peirce, Jonathan W. (2007). PsychoPy—Psychophysics software in Python. *J. Neurosci. Methods* 162:1, 8–13.

Peirce, Jonathan W. (2008). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* 2, p. 10.

Pitt, Mark A., L. Dilley, Keith Johnson, Scott Kiesling, William Raymond, E. Hume & E. Fosler-Lussier (2007). Buckeye corpus of conversational speech (2nd release). *Columbus, OH: Department* .

Revoile, S, J M Pickett, Lisa D. Holden & David Talkin (1982). Acoustic cues to final stop voicing for impaired-and normal hearing listeners. *J. Acoust. Soc. Am.* 72:4, 1145–1154.

Schiller, Niels O. (2020). Neurolinguistic approaches in morphology. *Oxford Research Encyclopedia, Linguistics* 1–23.

Sereno, Joan A. & Allard Jongman (1997). Processing of english inflectional morphology. *Mem. Cognit.* 25:4, 425–437.

Sharf, Donald J. (1962). Duration of Post-Stress intervocalic stops and preceding vowels. *Lang. Speech* 5:1, 26–30.

Twaddell, W. Freeman (1935). On defining the phoneme. *Language* 11:1, 5–62.

van der Walt, Stéfan, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu & scikit-image contributors (2014). Scikit-image: Image processing in Python. *PeerJ* 2, p. e453.

Zue, Victor W. & Martha Laferriere (1979). Acoustic study of medial /t, d/ in American English. *J. Acoust. Soc. Am.* 66:4, 1039–1050.