UNIVERSITY OF CALIFORNIA

SANTA CRUZ

REPORT ON THE REPLICATION OF SHINN & BLUMSTEIN (1984)

A thesis submitted in partial satisfaction
of the requirements for the degree of

MASTER OF ARTS

in

LINGUISTICS

by

John Callahan Boye-Lynn

August 2026

The thesis of John Callahan Boye-Lynn
is approved:

_____
Professor Amanda Rysling, Chair

_____
Professor Grant McGuire

_____
Professor Jaye Padgett

_____
Donald Smith
Interim Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

Abstract

Report on the Replication of Shinn & Blumstein (1984)

by

Cal Boye-Lynn

This study examines how American English listeners use two temporal parameters, amplitude rise time (ART) and transition duration (TD), to diagnose consonant manner, replicating previous work on the /b/-/w/ contrast (Shinn & Blumstein, 1984). A tripartite free response pattern is found, replicating stop and approximant responses at fast-ART, fast-TD, and slow-ART, slow-TD regions of the stimulus space, respectively, and replicating critical fricative responses in the slow-ART, fast-TD region of the stimulus space. This confirms that listeners do not require frication to report hearing fricatives. However, likely due to changes in synthesis parameters, we fail to replicate the original categorization responses, and find approximant responses in the fast-ART, slow-TD region of the stimulus space, akin to the results of Walsh & Diehl (1991), where the original experiment finds stop responses. Taken together, these results present a complex –if early– picture about the role of periodic information in manner perception.

1 Introduction

1.1 Consonant Manner and Temporal Parameters

Consonant manner is, perhaps, the least well-studied kind of contrast in speech perception. This is, historically, likely due to the well-validated assumption that changes in consonant manner are extremely perceptually obvious (Miller & Nicely, 1954, 1955; Steriade, 2008), which is typically linked to the large acoustic differences between phones of different manner classes. This assertion has, however, obscured a very simple question: how does a listener know they are hearing a consonant of a particular manner class? Or, put another way, how does a listener define what constitutes evidence of a particular consonant manner, and do the articulatorily-defined manner categories used throughout phonetics and phonology have well-defined perceptual analogues? This question has, with rare exception (see Section 1.4), gone almost entirely unanswered in research on speech perception.

Given the clear demonstrations in the literature that white noise masking, which destroys spectral information, most affects perception of consonant place (Miller & Nicely, 1954, 1955), it is reasonable to hypothesize that temporal parameters — those reflecting alterations in the timing of spectral and amplitudinal change — may be the most perceptually relevant for manner contrasts, as these parameters should be preserved under broad- and narrow-band simultaneous masking. In particular, we may expect that the duration of formant transitions and the duration of amplitude increases from silence to be relevant in manner perception, as previous acoustic studies clearly demonstrate that, at least in the American English labial series, stops demonstrate rapid changes in formant position and overall amplitude, where glides show slow changes in formant position and overall amplitude (Mack & Blumstein, 1983; Nittrouer & Studdert-Kennedy, 1986).

However, if listeners are to utilize this information, it must also be the case that (a) listeners are capable of representing formant position with sufficient temporal acuity to represent formant movement, and that this could be used across manner classes (including fricatives); (b) listeners are capable of representing intensity with sufficient temporal acuity to

render the relevant amplitudinal fluctuations, and that this could be used across manner classes; and (c) that listeners are willing to use both sets of information, alone or in conjunction, to make decisions about speech sounds' category. Many studies have been deployed which demonstrate that these conditions are met, and they are briefly reviewed below. Then, previous studies, including the subject of this replication, which directly test the relevancy of these parameters to manner perception, are reviewed, and major inconsistencies between them are highlighted.

1.2 Formant Movement in Speech Perception

First, we will need to establish that listeners care about formant contours, so that we know it is plausible a listener may employ differences in formant rate of change in the computation of some speech category. This section, then, will briefly review the relevant literature defining the formant, as an acoustic object, and discussing how alterations of formant position and contour are known to alter listener categorization.

To begin, let's briefly recall the articulatory origin of the formant. The vocal tract is traditionally divided into two components: sound sources (e.g., the vocal folds, turbulence-generation sites) and the filter (the nasal, oral and pharyngeal cavities) (Fant, 1960). During voicing, the vocal folds provide periodic perturbations in supraglottal pressure which can be spectrally decomposed into evenly spaced pure tones. These tones, the harmonics, provide widely distributed spectral energy which can be shaped by the vocal tract. Most harmonics will be damped, but some will be amplified by the vocal tract's resonating characteristics. When the articulators are held stationary, these resonances will remain relatively stable; however, when the articulators are in motion, the resonating properties of the oral tract change, causing changes in the resulting acoustics (Fant, 1960; Stevens, 1989). These resonances, the formants, have been shown to provide information about both consonant (Benkí, 2001; Cooper et al., 1952; Liberman et al., 1967; Nittrouer & Studdert-Kennedy, 1987; Nittrouer, 2002; Wagner et al., 2006; Shinn & Blumstein, 1984;

Sussman et al., 1991; Walsh & Diehl, 1991; i.a.) and vowel (Jenkins et al., 1983; Morrison, 2013; Nearey & Assmann, 1986; Nearey, 2012; Peterson & Barney, 1952; Strange, 2012; Strange et al., 1983; i.a.) categories in the perception of speech.

It is often implicitly assumed that the center frequency of the formant (often simply called "formant frequency") during the steady-state portion of a vowel is the primary, or perhaps only, piece of information used by a perceiver in the perception of vowels (Peterson & Barney, 1952; i.a.), and that a perceiver uses the onset of a formant transition and the beginning of its steady-state value to calculate an invariant "locus," from which the formant would have originated, as evidence of consonant place (Liberman et al., 1967; Sussman et al., 1991; Sussman et al., 1997). However, other reports demonstrate that listeners are concerned with more than just the static average of a formant's center frequency, or some idealized target, with portions of the formant contour –i.e., the amount each formant is changing– appearing relevant. This was first demonstrated by Strange et al. (1983), who found that CVC stimuli where the steady-state period of the vowel had been attenuated to silence still contained sufficient information for categorizing vowels, and thus that some steady-state target cannot act as an invariant cue (see also Jenkins et al. (1983), and see Strange & Jenkins (2012) for a review). Nearey & Assmann (1986) then demonstrated that the dynamic acoustic properties in the putative steady-state period of vowels in Canadian English also contribute to categorization (see also Morrison (2012) and Nearey (2012)). Together, these results seem to indicate that changes in formant center frequency may be just as useful to listeners as any particular value.

Fricative perception, by contrast, has long been argued to be largely insensitive to formant information by adulthood (Nittrouer & Studdert-Kennedy, 1987; Nittrouer, 2002), with limited exceptions in the transitions (Wagner et al., 2006). That said, very little work has examined what triggers a listener to categorize a consonant as a fricative, when other manner classes are possible, though Harris (1958) suggests that non-sibilant fricative categories may be evinced by periodic information, once a listener has diagnosed that the non-sibilant should be

categorized as a fricative. Jongman & McMurray (2011) also discuss several correlates which seem to impact fricative perception in currently unknown ways, which may also be recruited when distinguishing fricatives from non-fricatives.

This contrast between vowel and fricative perception has historically been associated with the distinction between periodic and aperiodic sounds. Periodic sounds are characterized by a regularly repeating fine structure, and are widely associated with voiced consonants and vowels due to the harmonic structure of the glottal waveform (Fant, 1960); where aperiodic sounds, characterized by a lack of regularly repeating fine structure, are widely associated with voiceless consonants, particularly fricatives. These two kinds of acoustic structure originate from very different sound sources: periodic structure is associated with vibration, where aperiodic structure is associated with turbulence caused by constrictions or obstructions. Subsequently, as vowels are dominated by periodic structure, where (particularly voiceless) fricatives are dominated by highly aperiodic structure, it has sometimes been assumed that the two do not share information (Ohala, 1981). If, as the argument would be made, fricatives are not only acoustically but perceptually distinguished largely or entirely by their frication, and vowels are not only acoustically but perceptually distinguished largely or entirely by the position of their formants, then the two should not meaningfully interact (i.e., signal elements typically taken as evidence that a fricative is present would never be used in the categorization of a vowel –or, non-fricative consonants, for that matter–, and vice versa). This is particularly obvious to gesturalist theories of speech perception, as different sound sources would, by necessity in a gesturalist theory, require differences in perceptual representation.

The broader notion that manner perception is relatively immutable has also been argued from studies of consonant confusions, where consonants of different manner and voicing are not typically confused for one another (Miller & Nicely, 1954, 1955). These classic studies asked listeners to transcribe speech in various kinds of noise, which allowed them to diagnose the frequency information which was important for listening to particular sounds,

and thus to predict what types of confusions different kinds and levels of noise would produce in the field. Importantly, however, this essentially removes only information about the frequencies present in the signal, but preserves its temporal dynamics. From this perspective, it is perhaps unsurprising that the fine acoustic information of the direction of a formant transition was lost, but the gross character of a sound's temporal dynamics (e.g., does it have a period of silence; does it have a sustained hiss or peak?) was retained. Interestingly, however, a more recent study which did not use white noise masking, Miller Willahan (2023), seemingly also demonstrated a more high-fidelity encoding of consonant manner.

Since this early work on speech-in-noise, multiple studies have, however, demonstrated that formant transition information is sometimes used to make a fricative distinction when the two fricatives contain highly similar aperiodic noise (Nittrouer & Studdert-Kennedy, 1987; Nittrouer, 2002; Wagner et al., 2006). This could imply, then, that formant transition information is included as a component of a perceiver's mental model of fricative category, at least for children listening to languages without highly similar fricatives and adults listening to languages with highly similar fricatives. Recall, then, that vowel category, which depends in large part on formant information, seems to be determined not by individual formant values, but instead by formant trajectories. Further, recall that the articulatory motions of approximants, fricatives and stops have one common acoustic effect: altering the resonances of the oral cavity. Given (a) formant information is used for fricative categorization (at least, between other fricatives), (b) some periodic information is to blame for non-sibilant fricative category decisions, (c) formant trajectory or contour is a component of vowel categorization, and (d) the formant trajectories of consonants at the same place but with different manners may be in a similar direction, but take place over different timescales, then it is perhaps possible that the rate of change of a formant may inform perception of consonant manner, perhaps even for fricatives.

1.3 Intensity Changes in Speech Perception

Second, we must establish that listeners can represent intensity fluctuation, and further that listeners use this information in the categorization of speech sounds. However, the role of stimulus intensity in speech perception is far less well-studied than the role of formant position. The mapping between the physical property of intensity and the psychophysical sensation of loudness is a primary endeavor of psychoacoustics (see Moore (2003) and Zwicker & Fastl (1999) for a review), and it has been demonstrated that listeners are capable of detecting amplitude modulation in a 1 kHz carrier of less than 5%, meaning listeners should certainly be able to detect all of the relevant gross amplitudinal fluctuations we will discuss (Schorer, 1986). But, in speech perception proper, very few studies have attempted to manipulate intensity to interrogate its status as a possible correlate, though some do exist. It has been established from studies on trading relations that the intensity of aspiration is relevant for VOT contrasts, and can enter a trading relationship with other subcomponents of the acoustics of aspiration (Kingston & Rysling 2023; Repp, 1979). Recent work has also demonstrated that intensity information contributes to spectral contrast effects (Viswanathan et al., 2009; Rysling & Kingston, unpublished). While these studies seem to indicate that intensity acts as a correlate for certain consonant distinctions, they do not establish whether it is the absolute intensity or the changes in intensity which were used by perceivers.

The work which has most closely examined intensity in manner perception comes from the literature on fricative distinctions. Harris (1958), for example, argues that listeners must care about the overall intensity, or distribution of the intensity, of frication when deciding on whether to treat a fricative as sibilant or non-sibilant. Additionally, overall intensity (and, the intensity of aperiodic energy within a particular frequency region) has a clear impact on fricative place perception, particularly for sibilants (Behrens & Blumstein, 1988; McMurray & Jongman, 2011). It has traditionally been argued that the amplitude difference between frication noise and the onset of either the F3 region (for sibilants) or the F5 region (for non-sibilants) determines fricative place categorization (Stevens, 1985); however, Hedrick &

Ohde (1993) demonstrate that these relative amplitude effects persist even when the following vowel is eliminated, pointing to local spectral or amplitudinal information within the frication noise itself providing the relevant evidence. We can thus safely conclude that absolute intensity is relevant to, minimally, distinguishing fricatives by sibilancy, and distinguishing among sibilant fricatives, though the role of amplitude change has not been clearly established, and nor is it clear how a listener diagnoses the presence of a fricative.

That said, other work explicitly argues against changes in intensity providing information about manner category. Kluender & Walsh (1992) constructed two pairs of continua between /ʃ/ and /t͡ʃ/ varying in either the duration of frication or the rate of intensity change from silence of that frication, with the non-varied correlate of each pair held constant, to disentangle which of the two correlates was more relevant to American English listeners. They found that, while varying frication duration resulted in a categorization function with a large span and a steeper slope around the presumed category boundary, varying the rate of intensity change resulted in a categorization function with a small span and a shallower, more linear slope. The authors took this to mean that it was duration, and not changes in intensity, which listeners use to make decisions. Notably, however, their sample size is relatively small by modern standards (though it was well-within contemporary standards, at n = 8 for their first experiment, with the intensity-change continuum pair, and n = 14 for their second experiment, with the duration continuum pair), and they describe having trained participants on the intensity-change continuum endpoints quite intensely, but provided no training for the duration continua. As it is now generally accepted that a linear categorization function can be caused by overtraining, or by the absence of top-down listening strategies which can be triggered by fatigue, it is unclear whether these results would replicate.

Regardless, it is clear listeners are capable of representing amplitudinal changes with sufficient acuity, and it is clear that listeners utilize absolute intensity in their decisions about some consonant contrasts, particularly contrasts among fricatives.

1.4 The American English /b/-/w/ Distinction

A version of each of these acoustic correlates, the rate of change of the formant transitions (hereafter “transition duration” or TD) and the rate of change of intensity (hereafter “amplitude rise time” or ART), have also been previously examined in two studies on the /b/-/w/ distinction in American English: Shinn & Blumstein (1984) and Walsh & Diehl (1991). Both studies put ART and TD into conflict in a series of continua and, in a categorization task, asked participants to label the resulting stimuli as “b” or “w.” Overall, their conclusions were completely different: Shinn & Blumstein (1984) found that ART is the more important of the two, largely determining response category, where Walsh & Diehl (1991) found that TD is the more important of the two, also largely determining response category. This represents a clear conflict.

In the earlier Shinn & Blumstein (1984), six continua, three with the vowel /a/ and three with the vowel /i/, were constructed. Each continuum varied in TD with different values of ART (see Section 2.1 for further details). In two continua, ART was held constant at either its fastest and most /b/-like value (the /b/-amplitude continuum) or at its slowest and most /w/-like value (the /w/-amplitude continuum). In the third continuum, ART covaried with TD (the standard continuum). Note that, while this put the two correlates in competition, it did not fully cross them, because ART was not varied independently of TD. As a result, we cannot tell whether an effect found just in the standard condition, where both vary, was due just to ART or due to an interaction between ART and TD.

These stimuli were then presented to listeners in four blocks alternating between categorization (called “forced-choice” in the original) and free response (called “free-identification” in the original), where the standard continuum was presented alone in the first two blocks, and the two non-standard continua (the /b/-amplitude and /w/-amplitude continua) were presented together in the second two blocks. This blocking structure has the potential to induce criterion shifts between blocks, as what constituted a good exemplar of a

token within a given block, and what correlates were relevant for understanding different

Fig. 1: Schematic of Shinn & Blumstein's (1984) Stimuli

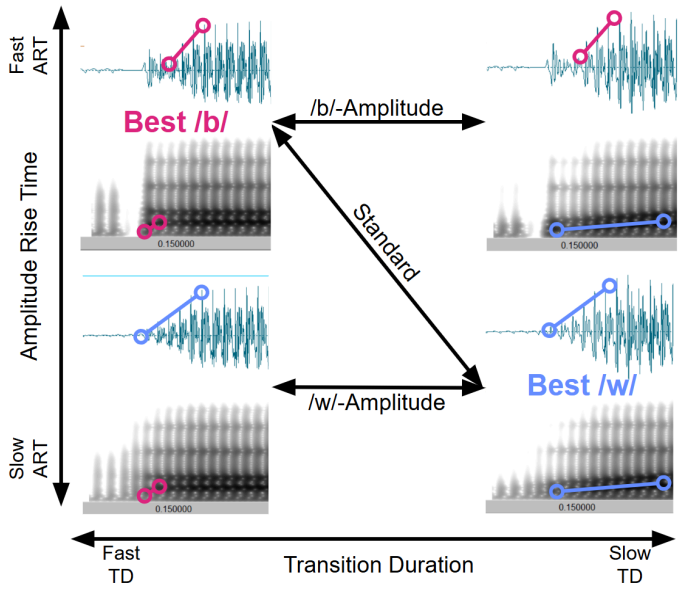


Figure 1: Schematic of the stimuli of Shinn & Blumstein (1984), depicting the waveforms and spectrograms of the replicated corner stimuli –i.e., those with extreme values of ART and TD. As indicated by the abscissa, the fastest and most /b/-like TDs are at left, where the slowest and most /w/-like are at right. This can also be seen in the rate of change of the formants in the spectrograms, as highlighted by pink (fast, /b/-like) or blue (slow, /w/-like) lines. As indicated by the ordinate, the fastest ARTs are at top, and slowest are at bottom. This can also be seen in the initial amplitude contour of the waveform. This then defines a parameter space, where a given stimulus is defined by its combination of ART and TD values. The lines between the corner stimuli represent the points in this space used to construct the continua. The horizontal lines represent those continua with a constant ART but varying TD (i.e., the /b/- and /w/-amplitude continua), and the diagonal represents the continuum which covaries the two (i.e., the standard).

regions of the acoustic space as distinct, changed (Brady & Darwin, 1978; Diehl et al., 1978; Eimas, 1963; Miller, 1953; Rosen, 1979). Thus, under an auditory category learning approach to speech perception, where it is recognized that what a listener considers relevant is decided to some extent on-line (Holt & Lotto, 2010), we might suspect that different blocks may have induced different categorization behavior.

In their categorization data, Shinn and Blumstein found that only the standard amplitude continuum (ART and TD covarying) resulted in a sigmoidal response curve with ceiling and

floor responses at the endpoints. In both of the fixed amplitude continua, response curves were more linear, and remained close to ceiling “b” responses (in the case of the /b/-amplitude continuum) or floor “b” responses (in the case of the /w/-amplitude continuum). From this, they concluded that ART must be the most relevant correlate to the distinction, as responses only change dramatically when ART varied (though, as noted above, this experiment cannot distinguish between whether it is variation in ART or variation in both ART and TD which mattered).

Their free response data complexified the story. While the most common responses were generally “b” and “w,” particular steps and particular continua had high rates of alternatives. Responses like “v” were quite common, particularly at the fast-TD region in the /w/-amplitude continuum and in the moderate-TD region of the standard continua. “d”-like responses were similarly common, particularly in the fast-TD region of the standard and /b/-amplitude continua. Overall, they found that stop-like responses dominated in their /b/-amplitude continuum and in the fast-TD region in their standard continuum, and that approximant-like responses dominated in the slow-TD portions of their standard and /w/-amplitude continua, but that fricative responses dominated in the moderate-TD portion of the standard continuum and the fast-TD portion of their /w/-amplitude continuum. They concluded that the presence of high frequency formant energy caused this fricative response pattern, in accordance with contemporary work on the role of gross spectral prominence of bursts in the categorization of coronal distinctions, cross-linguistically (Lahiri, 1984), and that ART ultimately acts as an invariant cue for the /b/-/w/ manner distinction. In other words, they argued that ART distinguished between /b/ and /w/, but that their stimuli contained a confound which indicated a non-bilabial place for a portion of the stimulus space, triggering “v” responses.

The idea that ART served as an invariant correlate of the American English /b/-/w/ distinction was challenged seven years later by Walsh & Diehl (1991), who found the opposite pattern in categorization alone. Walsh & Diehl (1991) crossed ART and TD in eight 11-step continua, delivered in a categorization task across four experiments. In their first and

third experiment, they created two continua varying in TD (20 - 120 ms, 10 ms steps) with two fixed ARTs (20 ms and 120 ms). In both experiments, the two continua were presented together in the same (mixed) categorization blocks. However, they differed in the detailed structure of their stimuli, with the first employing Klatt synthesized speech and the third employing sine-wave speech analogues. In their second and fourth experiment, they created two continua varying in ART (5 - 105 ms, 10 ms steps) with two fixed TDs (30 ms and 50 ms). Each continuum was then presented separately in its own block. The intermediate fixed TD values and separate blocking were chosen to make the ART changes more obvious (Diehl et al., 1978; Eimas, 1963), as pilot results demonstrated that extreme TDs overrode effects of ART. As in the first and third experiments, the second experiment contained Klatt synthesized speech, and the fourth experiment contained sine-wave analogues.

Fig. 2: Schematic of Walsh & Diehl's (1991) Stimuli

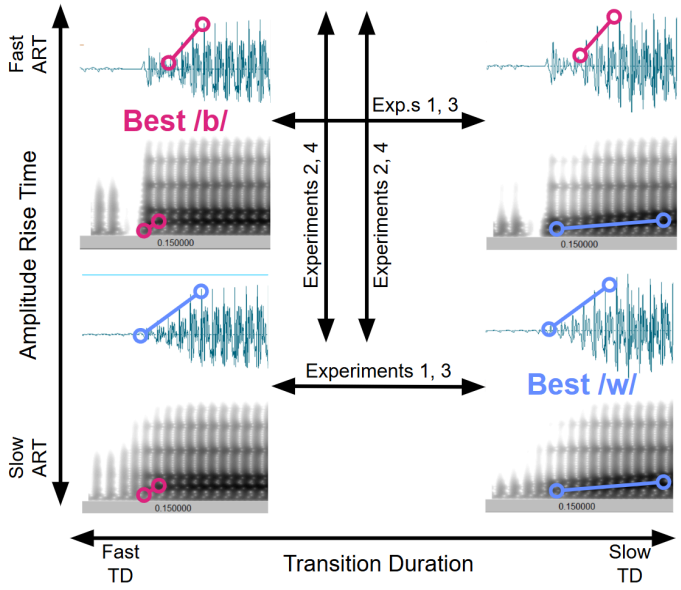


Figure 2: Using the same parameter space defined in Figure 1, with orthogonal variation in ART and TD, the stimuli of Walsh & Diehl (1991) are schematized. Their continua vary purely along TD, with two fixed TD values, analogous to the /b/- and /w/-amplitude continua of Shinn & Blumstein (1984) (Experiments 1 and 3); or vary purely along ART, with two fixed TD values occupying a far smaller acoustic range, and with no direct correspondence in Shinn & Blumstein (1984) (Experiments 2 and 4).

In both experiments varying in TD, responses were clearly sigmoidal, reaching ceiling and floor. However, in both experiments varying ART, responses were largely linear, with no continuum endpoint reaching more than 30% away from chance. While the curves in each experiment did seem to show some effects of the fixed correlate, Walsh and Diehl concluded that TD acts as the primary correlate of the distinction, and that these effects were due to low-level changes in the response of the auditory system.

In addition, Nittrouer & Studdert-Kennedy (1986) also conceptually replicated both Mack & Blumstein (1983) and Shinn & Blumstein (1984) in two experiments. In their production experiment, they found that, while /b/ had a larger amplitude slope than /w/, the “invariant” criteria Mack & Blumstein (1983) proposed to distinguish /b/ and /w/ productions did not successfully separate all of their two talkers’ /b/ productions from all of their talkers’ /w/ productions. Further, when the authors took typical /bV/ and /wV/ productions and interchanged their amplitude contours, listeners in a free response task consistently describe the resulting stimuli as most closely resembling the original sound (e.g., a /bV/ token with a /w/-like amplitude contour was described most often as /b/-like). Thus, a major change in ART was insufficient to alter the final decision, which contrasts sharply with Shinn & Blumstein’s (1984) claim that ART acts as an invariant manner cue.

Together these form a bizarre picture. Shinn & Blumstein (1984) showed that ART seems to matter for making the /b/-/w/ distinction (though their work could not disentangle its effects from TD) and demonstrated that fricative responses could sometimes occur within the relevant stimulus space. Note, further, that the categorization and free response results of their experiment seem to show a similar kind of boundary between stop-like responses and non-stop-like responses. Walsh & Diehl (1991), meanwhile, demonstrated that TD dominated response patterns when fully crossed with ART, and Nittrouer & Studdert-Kennedy (1986) provide further evidence that ART changes alone cannot override other elements of the signal. Subsequently, while it is clear that listeners can and do use ART- and TD-like correlates, it is unclear what elements of the signal American English listeners are using to

diagnose consonant manner. The present study, then, aims to clarify the role of ART and TD in the American English /b/-/w/ distinction by replicating Shinn & Blumstein (1984), with the goal of integrating the seemingly incongruous results of these studies towards a unified understanding of the effects of periodic information on the categorization of consonant manner.

2 Methods

2.1 Stimuli

Six continua varying linearly in TD from 25 ms (the most /b/-like) to 75 ms (the most /w/-like) in eleven 5 ms steps comprised the stimuli of the original paper, three continua in each vowel condition (vowel was a between-subjects factor). Of those three, one, the standard amplitude continuum, also varied in ART from 5 ms (the most /b/-like) to 55 ms (the most /w/-like) such that ART and TD covaried. In the other two continua, ART was held constant at either 5 ms (the most /b/-like it could be) or 55 ms (the most /w/-like it could be). Note that, within a given vowel condition, the first stimulus of the standard continuum was identical to the first stimulus of the /b/-amplitude continuum, and the last stimulus of the standard continuum was identical to the last stimulus of the /w/-amplitude continuum. So, two stimuli, the standard endpoints, which should have sounded the most like /b/ and /w/, appeared twice as frequently as any other stimulus.

The original stimuli were created using a Klatt synthesizer, and most details for the endpoint stimuli are available in the original paper's Table 1. However, details about the bandwidth of F1, F4 and F5, and the center frequency of F4 and F5, are not provided. The values used for the present study are included in the materials linked below, along with the R (R Core Team, 2023) and Praat (Boersma & Weenink, 2023) code used to construct the stimuli. Note that, unlike the original paper, which used the synthesizer in its parallel configuration, allowing for the independent specification of amplitude and bandwidth characteristics for each formant, the replication used the synthesizer in its cascade

configuration[1], which does not allow for the independent specification of amplitude, though does permits specification of bandwidth. This results in stimulus amplitude largely depending on the frequency of F1, such that amplitude contour is not held entirely constant across conditions where the AV (amplitude of voicing) parameter is fixed (i.e., the /b/- and /w/-amplitude continua). The stimuli also contain less intense, lower-bandwidth higher formants, and have a phenomenally “softer” quality.

The center frequency values for F4 and F5 were held constant at 3.3 and 3.8 kHz, respectively. While their constancy is certainly unnatural, these values were drawn from Walsh & Diehl (1991) whose stimuli, though varying from the present stimuli in many ways, were similarly informed by Mack & Blumstein (1983). Subsequently, these higher formant values, sometimes implicated in talker recognition (Barreda, 2016; Baumann & Belin, 2008), should reflect someone with appropriate facial morphology. Additionally, while F4 and F5 can vary depending on the lower resonances of the vocal tract (Fant, 1960; Stevens, 1989), no studies known to the author have demonstrated their use in perception of vowel or consonant category. Their frequency should thus have at worst minor impacts on consonant categorization.

The bandwidth values for F1, F4 and F5 (hereafter B1, B4 and B5) were set at 60, 230 and 290 Hz, respectively. While the higher bandwidth values are less likely to matter for consonant or vowel contrasts, B1, which could have an effect, was chosen to align with both the pattern of stated values provided in the original paper and those provided by Klatt (1980)[2], who listed 60, 110 and 130 Hz as the B1, B2 and B3 for /b/ and 50, 80 and 60 Hz as the B1, B2 and B3 of /w/ (Klatt, 1980). Notice that the values of B2 and B3 used in Shinn & Blumstein’s (1984) original stimuli (110 and 170 Hz) are higher than those provided by Klatt

[1] This may impact fast-ART, slow-TD stimuli in a manner not fully understood. To multiple phonetically trained listeners, the cascade version of the stimuli were more approximant-like, where the parallel version were more stop-like. An extension will verify this pattern in naive listeners. See also Section 3.2.

[2] Klatt (1980) provides no complete rationale for these, but they have been traditionally used as standard Klatt synthesizer parameters, particularly contemporarily to the original study.

(1980), so we have also selected a B1 which is somewhat higher than that suggested by Klatt (1980). After the difference between B1 and B2 (50 Hz), we increase every successive B# by 60 Hz, mimicking the difference between the original B2 and B3. This follows the general trend that higher formants have larger bandwidths (though they do not always, see Klatt,1980, and Kent & Vorperian, 2018). Note that, due to an error, the B1 used in the present study narrows after the initial 50 ms of simulated prevoicing, unlike in the original, where B1 is held constant. Note that all stimuli had simulated prevoicing, though only stimuli with a /b/-like amplitude contour had a period of silence between the simulated prevoicing and the onset of the rest of the consonant.

The endpoint stimuli settings were arranged manually, with blank values interpolated by hand with the assistance of a Python script. This was then converted into a CSV file and passed to an R (R Core Team, 2023) script which linearly interpolated all parameters at each timepoint, creating eleven total sets of KlattGrid settings. Praat (Boersma & Weenink, 2023) was then used to generate the stimuli as a series of uncompressed WAV files at a sampling rate of 44100 Hz.

2.2 Tasks

Participants were asked to complete two tasks during the experiment, as in the original: categorization and free response. Tasks were only ever performed in dedicated blocks (i.e., participants were not asked to respond in multiple ways on a given trial or within a given block) and participants were provided explicit instructions on the structure of each task before the experiment began and before each block. For every categorization block with a certain set of stimuli, a free response block with the same set of stimuli always followed, creating pairs of categorization and free response blocks.

During a categorization trial, participants listened to one of the stimuli over headphones (Sennheiser HD 280 PRO) and, once they had made a decision about the sound they heard, were asked to press a button on a wired keyboard to label the sound as “b” or “w.” At the

onset of each trial, a fixation cross was presented in the center of the screen for 500 ms. It then disappeared and the response options were provided on screen as the stimulus began playing. <B> and <f> were presented on the left side of the screen, reminding participants that the 'f' key represented the "b" label, and <W> and <j> were presented on the right side of the screen, reminding them that the 'j' key represented the "w" label[3]. Stimuli were each 445 ms long, with PsychoPy (Pierce et al., 2019) providing a 1 ms raised cosine taper at the onset and offset, though the PsychoPy routine was set to allow for up to 500 ms of playtime. Participants could provide a response any time after the onset of the sound, including during the 500 ms window allotted for the sound itself and a silent 1500 ms window which followed. Once the participant had provided a response and the sound window had ended (i.e., after 1000 ms from the onset of the trial), the trial would end and the next trial would begin.[4] Alternatively, a trial would end automatically if the participant failed to respond at any point during the entire 2000 ms response window. There was no additional intertrial interval.

The structure of a free response trial was highly similar. Participants would first be presented a fixation cross for 500 ms, then the stimulus would begin. As the stimulus began, the fixation cross would disappear from the screen and a textbox would appear, in which the participant could type a response. Participants could then press the 'space' key to submit their response at any point after the onset of the stimulus. Once a participant had entered a response and the 500 ms sound window had finished, they would move on to the next trial. Participants then had 4000 ms after the offset of the sound window in which they could still respond. If they did not respond in time, the experiment would automatically progress to the next trial.

[3] Note that handedness could play a role in response time or response accuracy. We only examine the response category, so we expect effects of handedness to be small. If present, they would be "w"-biasing.

[4] Waiting until the end of the sound window before proceeding to the next trial was done to avoid an audio handling issue in PsychoPy v.2023.2.3., where early responses, which ended the sound window prematurely, would prevent the sound in the next trial from playing.

At the end of the experiment, participants were asked to complete a short debriefing and demographics questionnaire, to gather information about their experience with the experiment and their language history. The forms provided to participants are available here < S&B Replication >. Stimuli, experiment files and all other materials are available here: < S&B Replication >.

2.3 Blocking

All versions of the experiment involved pairs of blocks which alternated in task, with categorization in the first block of the pair and free response in the second. Within a given pair of blocks, the stimulus set was always identical (i.e., if you heard a given set of stimuli in a categorization block, you would then be hearing that same set of stimuli in the following free response block). The arrangement of blocks was manipulated as a between-subjects factor, with three conditions: Blocked (mimicking the original experiment), Unblocked (mimicking a preceding pilot study[5]) and Partial (an intermediary). Regardless of blocking condition, a given experiment always had 1320 trials, with 18 test trials per condition.

In the fully Blocked condition, participants saw four total blocks (two categorization-free response pairs), mimicking the structure of the original experiment, where the standard and non-standard continua were presented separately. In the first pair, participants only heard stimuli from the 11-stimulus standard amplitude continuum. Participants first attempted the task while listening to five pre-selected stimuli from the standard continuum in a set order, consisting of the endpoints followed by three increasingly intermediate stimuli. Explicit feedback was provided for the first two stimuli. Then, participants began a “burn-in” period which included two full repetitions of the stimulus set (22 trials per block) presented in pseudorandom order such that the entire stimulus set is played before a token is repeated.

[5] This pilot study was run by myself and Emily Knick as part of a project in Phonetics A at UCSC, in Spring 2024, under the supervision of Grant McGuire. Thank you, Emily, for your help in getting this project off the ground.

After the burn-in period, participants listened to eighteen repetitions of the stimulus set (198 per block), again in pseudorandom order, resulting in a total of 440 trials in the first pair of blocks. In the second pair, participants only heard stimuli from the non-standard continua (i.e., the /b/- and /w/-amplitude continua). Again, they heard five pre-selected stimuli (two non-standard endpoints and three intermediary) to help familiarize them with the task, then were given two full repetitions of the stimulus set as burn-in (i.e., 44 per block) before they accessed the remainder (396 per block), resulting in a total of 880 trials in the second pair of blocks. Participants were given breaks at the end of each block.

In the Unblocked condition, participants saw only two total blocks (one pair), mimicking the blocking structure of a previous pilot. In both blocks, participants listened to all three continua together, and the structure of each block was identical. Five pre-selected stimuli, identical to those used in the first two blocks in the Blocked condition (i.e., drawn solely from the standard continuum), were used for familiarization, followed by a burn-in period consisting of two full repetitions of the stimulus set (66 per block) and the test trials consisting of eighteen repetitions of the same stimulus set (594 per block). Participants in this condition were given one break in the middle of the categorization block, two breaks at the third and two-thirds mark of the free response block, and one between the categorization and free response blocks.

Finally, in the Partial condition, participants were presented with four total blocks (two pairs). The number of trials in each block was identical to the fully Blocked condition, but the entire stimulus set, including all three continua, were presented together in all four blocks, as in the Unblocked condition. To maintain the same number of trials in the burn-in periods of each block, a subset of the entire stimulus set was chosen to act as a representative sample. Eleven stimuli (/b/-amplitude steps 5, 8 and 11; standard steps 1, 4, 6, 8 and 11; /w/-amplitude steps 1, 4, 7) were chosen to act as the burn-in stimuli for the first two blocks, and eleven additional stimuli (/b/-amplitude 1, 3, 7, 9; standard 3, 5, 7, 9; /w/-amplitude 3, 5, 9, 11) were added in the second two blocks. These were chosen such that (a) the endpoints

of each continuum were always included, (b) the widest distribution of stimuli from across the continuum were included and (c) a minimum number of adjacent stimuli were included. This should provide listeners with a good representation of the full stimulus space while keeping the number of training, burn-in and test trials identical to the Blocked condition.

2.4 Participants

So far, 162 members of the University of California, Santa Cruz community, largely students from undergraduate linguistics courses, have participated. All provided informed consent and participated for course credit or $12.00 USD in compensation. Participants who self-described as having non-normative hearing, dyslexia, auditory processing difficulties or a speech delay; who began learning American English after six; who did not live in the United States before they were six; or who chose not to provide information about their language background were excluded. A total of 39 participants were excluded, 37 due to the aforementioned criteria, and two due to technical errors.

The remaining 123 participants are evenly divided across conditions, though we will be examining only data from the first 120 (20/cell; age 20.05 ± 2.03). This is double the number of participants per between-subject condition from the original study's, which had 10 per cell, but is still below our heuristic target of 30 per cell. Of these 120 participants, 32 self-reported that they were neurodivergent or that they had a disability affecting their performance in school. Ten self-reported being left-handed. Of the included participants, most had been exposed to at least one other language, with Spanish (86) being the most common. Participants are still being recruited to reach a target of 180 (i.e., 30 per cell).

3 Results[6]

3.1 Categorization

Overall, the continua in all conditions are categorized similarly, although there are important exceptions. The results of the categorization task, separated by blocking and vowel conditions, are depicted below in Fig. 3. Continuum step is on the abscissa, with 1 representing the fastest TD, and 11 representing the slowest; and, proportion of “b” responses is on the ordinate. Each between-subjects condition (i.e., combination of blocking, vowel) is displayed in a separate panel. The first row represents the /a/ conditions and the second row the /i/ conditions; meanwhile the leftmost column represents the fully Blocked conditions, the middle-left column represents the Partial conditions and the middle-right column represents the Unblocked conditions. The original results, as estimated from the Figure 3 of Shinn & Blumstein (1984), are presented in the rightmost column. The two leftmost panels should represent the closest replication of the original study; and, thus, the leftmost and rightmost panels should be the most similar. Within each panel, the red circles represent responses to the /b/-amplitude continuum, the yellow triangles represent responses to the standard continuum, and the blue squares represent responses to the /w/-amplitude continuum. Note that the standard continuum has, in all cases, a relatively high span (though it never reaches 100% “b” responses, indicating listeners are never certain they are hearing a /b/), which looks similar to the original. Additionally, the /w/-amplitude continua visually show some flattening like the original. Finally, the /b/-amplitude continua looks nothing like the original, though does show some flattening in the Blocked conditions. These patterns are all statistically credible, as discussed in detail below.

[6] While I will be fitting a logistic regression model to the categorization data below, and using a similar multinomial regression for the free response data, these should be taken as a demonstration of the analytical approach which will be taken once data collection is complete.

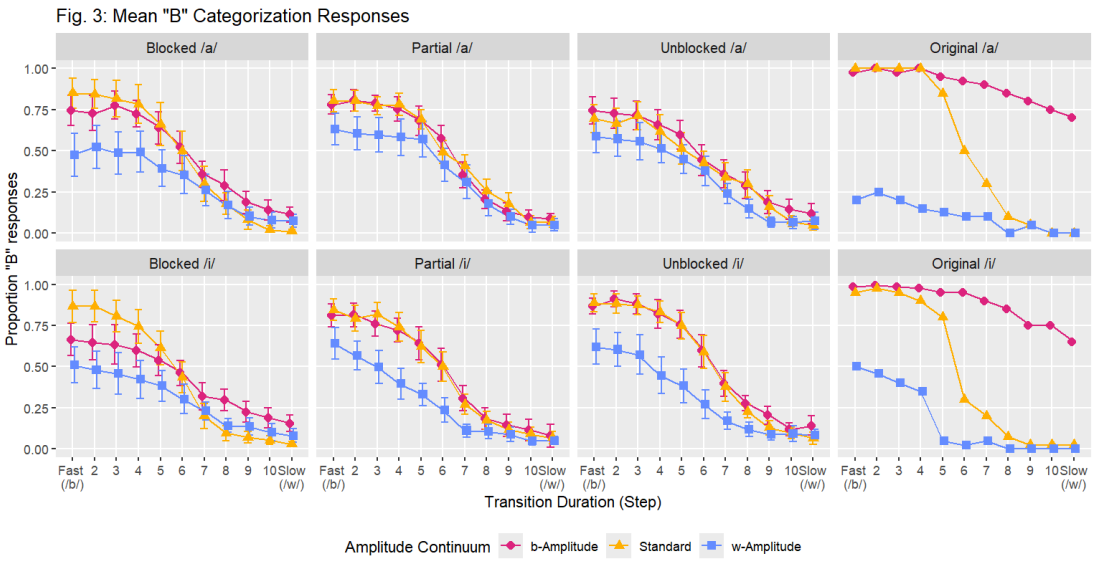


Figure 3: Proportion “B” responses split by between-subjects condition (i.e., separated by blocking and vowel condition), with reconstructions of the original responses in Shinn & Blumstein (1984), via visual estimation of the original figures, for comparison. Within each panel, the abscissa is TD, with the fastest and most /b/-like at left, and the slowest and most /w/-like at right. The ordinate is proportion “B” responses. The lines represent the continua: red circles represent responses to the /b/-amplitude continuum, yellow triangles represent responses to the standard continuum, and blue squares represent responses to the /w/-amplitude continua. Intervals are 95% confidence (note, not credible) intervals.

Six Bayesian logistic mixed-effects models were fit to the response data, one for each of the between-subjects conditions, using the *brms* (Bürkner, 2017) package in R (R Core Team, 2023). Note that “b” responses were coded as ones and “w” responses as zeroes. In each model, TD step and continuum (standard, /b/-amplitude or /w/-amplitude) were coded as fixed effects, with step centered and scaled, and continuum treatment coded such that the standard continuum acted as the baseline. Additionally, pilot results demonstrated that responses to the fastest TD steps of certain continua varied wildly between individuals, which coincides with the region of stimulus space where Shinn & Blumstein (1984) note an uptick in “v” responses in their free response task. This could be because participants were consistently concluding they heard a third category (likely /v/) and were sorting it consciously into one of the two provided labels. Participants were thus asked in the debriefing survey "Is 'v' more like 'b' or 'w', to you, if you had to choose?" This “v”-labeling preference is also

included as a fixed effect with treatment coding such that the response “B” was considered the baseline. Note that this is unbalanced both within and across conditions (i.e., we did not control for “v”-labeling preference). Random effects of participant, with intercept and slopes for step and continuum were also included.[7]

In the Blocked /a/ condition (n = 20), there is only a clear main effect of step (β = -3.07, Est.Error = 0.39, CrI: [-3.87, -2.34]), with neither the posterior for the effect of /b/-amplitude continuum (β = -0.10, Est.Error = 0.32, CrI: [-0.73, 0.54]) nor the posterior for the effect of /w/-amplitude (β = -0.56, Est.Error = 0.49, CrI: [-1.54, 0.43]) continuum, relative to the standard amplitude continuum, credibly different from zero. This reflects that all continua show reduced “w” responses as TD step increases, but they do not seem to differ in where the boundaries between “b” and “w” responses lay. Notably, both the interaction between step and /b/-amplitude (β = 1.46, Est.Error = 0.33, CrI: [0.85, 2.12]) and the interaction between step and /w/-amplitude (β = 1.66, Est.Error = 0.33, CrI: [1.02, 2.31]) are credibly positive. This means that these continua credibly change less as a consequence of changes in step, reflecting the blunting or flattening of the /b/- and /w/-amplitude continua visible in the graph. Interestingly, however, neither the main effect of “v”-labeling preference (β = -0.96, Est.Error = 0.64, CrI: [-2.28, 0.33]), nor any of its two-way interactions (Step x Preference: β = 0.60, Est.Error = 0.65, CrI: [-0.65, 1.85]; /b/-amplitude Continuum x Preference: β = 0.68, Est.Error = 0.55, CrI: [-0.44, 1.76]; /w/-amplitude Continuum x Preference: β = -1.07, Est.Error = 0.83, CrI: [-2.73, 0.56]), are credibly non-zero. This implies that preference alone may not change either the way in which a longer TD depresses “b” responses, or when that depression begins. Additionally, neither three-way interaction is credibly non-zero (Step x /b/-amplitude Continuum x Preference: β = -0.50, Est.Error = 0.54, CrI: [-1.64, 0.55]; Step x /w/-amplitude Continuum x Preference: β = 0.22, Est.Error = 0.54, CrI: [-0.84, 1.27]). This then implies that

[7] Including “v”-labeling in the random effect structure introduced instability, leading to R-hat values above the 1.01 standard suggested by Vehtari et al. (2021). So, this random effect structure was used instead.

“v”-labeling preference did not change the shape of either function, relative to the standard amplitude continuum.

Similar results are obtained for the Blocked /i/ condition (n = 20). There is a credibly negative effect of step (β = -2.43, Est.Error = 0.21, CrI: [-2.86, -2.04]) and credibly positive interactions between step and /b/-amplitude (β = 1.40, Est.Error = 0.21, CrI: [0.99, 1.81]), and step and /w/-amplitude (β = 1.46, Est.Error = 0.19, CrI: [1.09, 1.86]), but neither effect of continuum is credibly non-zero (/b/-amplitude Continuum: β = 0.23, Est.Error = 0.25, CrI: [-0.26, 0.74]; /w/-amplitude Continuum: β = -0.57, Est.Error = 0.32, CrI: [-1.22, 0.05]). This is again in line with the shape of the standard continuum being credibly different from the /b/- and /w/-amplitude continua, without changing its crossover point very much, reflecting the flattening visible in the figure. Like the Blocked /a/ condition, preference is not credibly non-zero (β = -0.77, Est.Error = 0.69, CrI: [-2.14, 0.61]), and neither are any of its interactions.

Overall, in the Blocked conditions, there are clear interactions between step and the non-standard continua without any main effect of continuum, seemingly indicating that, while the boundary between “b” and “w” responses does not change, the steepness of the function is blunted in the non-standard continua. This seems to imply that participants may have been less convinced they were hearing a /b/ or /w/, or could indicate overtraining. Participants notably demonstrated no difference in category boundary on the basis of their preference for labeling “v” as “b” or “w”. This could imply that few participants consistently heard “v”-like stimuli, or could imply that the boundary between a “w” response and a non-“w” response was what mattered to participants overall, and that this boundary remained consistent across continua. Note that, while both non-standard continua show a reduced span in both Blocked conditions, responses are centered around chance, and are not pinned at ceiling or floor, as Shinn & Blumstein (1984) found. Thus, we do not replicate their categorization responses, even in the condition with the most similar blocking structure.

In the Partial /a/ condition (n = 20), there is again a credibly negative effect of step (β = -1.68, Est.Error = 0.13, CrI: [-1.93, -1.42]) and no credible effect of continuum (/b/-amplitude Continuum: β = -0.04, Est.Error = 0.13, CrI: [-0.29, 0.21]; /w/-amplitude Continuum: β = -0.27, Est.Error = 0.25, CrI: [-0.77, 0.23]). Unlike in the Blocked conditions, however, the interactions between step and the non-standard continua are not credibly non-zero (Step x /b/-amplitude Continuum: β = 0.02, Est.Error = 0.10, CrI: [-0.17, 0.21]; Step x /w/-amplitude Continuum: β = 0.18, Est.Error = 0.14, CrI: [-0.11, 0.46]), reflecting the consistent shape between continua. Collectively, this implies that all continua show reduced “b” responses as step increases (though this change is potentially less intense than the changes in the Blocked conditions, given the lower estimate of the effect of step), but that there are no major changes in crossover point or function shape by continuum. This is reflected visually: the three continua trace a highly similar path. Notice, however, that the /w/-amplitude continuum is still numerically lower than the other two continua, which is common across between-subjects conditions for the faster-TD end of the continua. The main effect of “v”-labeling preference and all interactions between preference and other predictors are not credibly non-zero, though the interaction between step and preference is so only narrowly (β = 0.35, Est.Error = 0.22, CrI: [-0.09, 1.79]). This means that, overall, “v”-labeling preference does not predict changes in the crossover points or shapes of participant categorization functions, but, if it predicts anything, it would be that “w”-labelers may be slower to change their answer as stimulus step changes.

In the Partial /i/ condition (n = 20), we again find a credibly negative effect of step (β = -1.87, Est.Error = 0.23, CrI: [-2.34, -1.42]) and no credible effect of the /b/-amplitude continuum (β = 0.10, Est.Error = 0.12, CrI: [-0.13, 0.33]). However, unlike the previous conditions, we see a credibly negative effect of the /w/-amplitude continuum (β = -1.33, Est.Error = 0.27, CrI: [-1.66, -0.62]). This implies that the predicted crossover point for the /w/-amplitude continuum is earlier, by a fair margin, than those of the other two continua, which pattern more closely. This is reflected visually: the /b/- and standard amplitude continua

are almost on top of each other, where the /w/-amplitude continuum traces roughly the same shape, but appears to descend earlier. The interaction between step and /b/-amplitude (β = 0.16, Est.Error = 0.12, CrI: [-0.09, 0.38]) is not credibly non-zero, but the interaction between step and /w/-amplitude (β = 0.36, Est.Error = 0.14, CrI: [0.09, 0.65]) is credibly positive. This means the /w/-amplitude continuum shows a small decrement in the rate at which listeners change their answer, relative to the standard continuum, reflecting the reduced span. As before, none of the effects or interactions related to “v”-labeling preference are credibly non-zero, indicating that no responses were better predicted on the basis of a participant’s “v”-labeling preference.

Overall, the Partial conditions pattern similarly, with little of the flattening in the /b/-amplitude continua which characterized the Blocked conditions. Generally, all continua trace a similar path, though the /w/-amplitude continuum is shifted downwards or leftwards in the /i/ condition (and is numerically lower in the /a/ condition, though this is not reflected in either the effect of /w/-amplitude continuum or its interaction with step), such that fewer “b” responses are elicited across the board, and they drop off more slowly as step increases. Note, however, that the standard continuum has a lower span, due to a lower fast-TD endpoint response, than appears in the Blocked conditions. This could mean that these stimuli were never particularly convincing /b/’s, or could perhaps reflect some exhaustion on the part of participants. The latter could also explain the smaller estimate of the effect of step. Again, this does not resemble the original study, with the /b/-amplitude continuum in particular showing no signs of flattening.

In the Unblocked /a/ condition (n = 20), the effect of step is credibly negative (β = -1.52, Est.Error = 0.21, CrI: [-1.95, -1.11]) but the effects of /b/-amplitude (β = 0.15, Est.Error = 0.10, CrI: [-0.05, 0.34]) and /w/-amplitude (β = -0.29, Est.Error = 0.29, CrI: [-0.86, 0.28]) are not credibly non-zero. Though, numerically, the /b/-amplitude continuum appears shifted upwards or rightward from the standard continuum, which itself appears shifted upwards or rightward from the /w/-amplitude continuum; this is reflected in the estimated beta values. Importantly,

the step and /b/-amplitude continuum interaction (β = -0.06, Est.Error = 0.13, CrI: [-0.31, 0.21]) and the step and /w/-amplitude interaction (β = -0.01, Est.Error = 0.14, CrI: [-0.29, 0.27]) are not credibly non-zero, with tiny estimated effects relative to the estimated error. As such, it seems that, for these continua, we find no major differences in crossover point or in overall function shape. Finally, none of the posteriors for effects or interactions related to "v"-labeling preference are credibly non-zero.

Finally, in the Unblocked /i/ results, we again find a credibly negative effect of step (β = -2.00, Est.Error = 0.30, CrI: [-2.50, -1.52]), but the effect of the /b/-amplitude continuum (β = -0.04, Est.Error = 0.26, CrI: [-0.55, 0.47]) is not credibly non-zero and the effect of the /w/-amplitude continuum (β = -0.69, Est.Error = 0.34, CrI: [-1.36, 0.00]) is just credibly negative. Overall, this means there are no major changes in crossover point, though the /w/-amplitude continuum may have a boundary closer to the faster-TD end of the continuum. Neither interaction between step and continuum, however, are credibly non-zero (Step x /b/-amplitude Continuum: β = 0.13, Est.Error = 0.15, CrI: [-0.17, 0.44]; Step x /w/-amplitude Continuum: β = 0.22, Est.Error = 0.29, CrI: [-0.34, 0.79]), implying the continua do not strongly differ in shape. This seems somewhat at odds with the clear visual difference between the /w/-amplitude continuum and the other two continua. One potential explanation lies in the "v"-labeling preference predictors. While the effect of "v"-labeling is not credibly non-zero (β = -0.05, Est.Error = 0.45, CrI: [-0.92, 0.83]) and the interaction between "v"-labeling preference and step is also not credibly non-zero, the two-way interaction between /w/-amplitude and "v"-labeling preference is credibly negative (β = -0.95, Est.Error = 0.43, CrI: [-1.83, -0.14]), implying that the crossover point for "w"-labelers is earlier in the /w/-amplitude continuum. Finally, the three-way interaction between step, /w/-amplitude continuum and "v"-labeling preference is credibly positive (β = 0.79, Est.Error = 0.36, CrI: [0.08, 1.51]), reflecting a flattening or blunting for "w"-labelers along the /w/-amplitude continuum.

Overall, then, we generally find that the Unblocked conditions demonstrate relatively similar categorization functions across continua, though the /w/-amplitude continuum of the /i/ condition seems to demonstrate a difference in crossover point and a strong divergence between “b”- and “w”-labeler responses, such that those who prefer to label /v/-like sounds as “w” have a far earlier crossover point. This numerical pattern is visible in most conditions (i.e., the solid line with blue squares is above the dotted line with blue squares in Figure 4, below, for most conditions), but only in the Unblocked /i/ condition are confidence intervals for most fast-TD steps clearly non-overlapping. In most other conditions, the confidence intervals are overlapped; though, note that, while visually obvious in the Blocked /a/ condition, no effect or interaction of “v”-labeling is credibly non-zero. Regardless, this “v”-labeling-depenedent pattern may contribute to the apparent difference between the /w/-amplitude continuum and the other continua in the Unblocked /i/ condition. Otherwise, however, these functions are of a consistent shape.

Fig. 4: Mean "B" Categorization Responses, By V-Labeler Preference

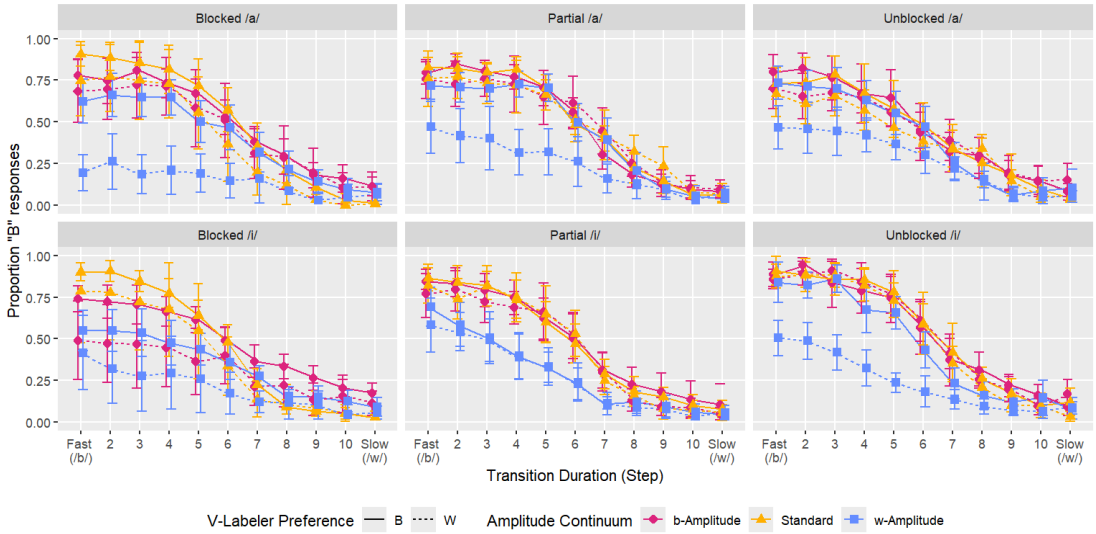


Figure 4: Proportion “B” responses broken out by listeners’ “v”-labeling preference. Notice that the /b/-amplitude continuum and the standard continuum generally do not separate by preference, where the /w/-amplitude continuum does separate by preference.

To sum up: we do not replicate the categorization responses of Shinn & Blumstein (1984), though some of the original patterns remain. As in the original study, the standard continuum

clearly elicits largely “b” responses from the fast-TD end of the continuum, and largely “w” responses from the slow-TD end of the continuum, though the low maximum value in some conditions demonstrates that listeners were not fully convinced they were hearing something /b/-like. Additionally, the /w/-amplitude response pattern found in the original study’s /a/ condition is replicated in our Blocked /a/, Blocked /i/, Partial /i/ and Unblocked /i/ conditions, with a moderately depressed /w/-amplitude continuum. This could reflect the presence of a third category. Finally, however, the behavior of the /b/-amplitude continuum does not match the original results at all. While it is depressed in the Blocked conditions, with a lower span, it does not display the near-ceiling results found by the original authors. Then, in the Non-Blocked conditions (Partial and Unblocked), the /b/-amplitude continuum entirely overlaps with the standard continuum. This strongly implies that the flattening we find in the Blocked conditions is a range effect of some kind, though it is not nearly large enough to account for the original results. This could be due to a number of factors, including the slightly different construction of the stimuli, idiosyncrasies of the participants, or changes in American English which have come about during the time lag between the original and present study.

3.2 Free Response

In the free response task, participants were asked to type out whatever best represented the sound they heard. Many indicated that they heard sounds which did not fit the labels “b” or “w”, either during the whole duration of the experiment or specifically during the free response portions. However, some participants indicated that they did not realize they were allowed to use answers other than “b” or “w” during the free response blocks, or answers other than “b”, “w” or “v” (“va” or “vi”, depending on vowel condition, was used as an example during the in-experiment instructions as a demonstration of the kind of previously unused label one could use during free response, if need be). Consequently, the counts of “b” and “w” responses are likely overcounts, and the counts of other responses (with the possible exception of “v”) are likely undercounts. Later participants were told verbally that they may

hear other sounds throughout, in addition to written instructions which appeared on screen during the experiment.

Responses were coded by first eliminating spaces at the beginning of the string participants had typed. As the space bar was used to trigger the end of a trial, participants occasionally pressed space at the start of a trial accidentally. Responses were binned based on their first letter. Some responses included multiple consonant letters (e.g., “cr”, “gl”, “bw”), but these are largely ignored for this analysis. Additionally, some participants used the letter “c” as the initial consonant in their responses. Participants were asked to list any letters which could have been ambiguous in the debriefing survey and, where participants provided an explanation, “c” responses were recoded as either “k” responses (corresponding to a “hard” c) or “s” responses (corresponding to a “soft” c). When participants did not provide an explanation, these responses were coded as neither “k” nor “s”, unless the participant used a word containing one or the other (e.g., the response “choir” would be recoded as “k”-initial). Responses were then binned into the manner categories stop, fricative and approximant based on the initial letter, excluding non-responses, vowel-initial responses and already-ambiguous responses. Clusters were then coded in the case of an initial letter which indicated a stop and a peninitial letter which indicated an approximant, though only the initial letter’s associated manner was submitted to the statistical analyses described below.

Heatmaps for each manner class (approximants, Fig. 5; stops, Fig. 6; and fricatives, Fig. 7) are below. Each heatmap contains eight panels, one for each of the between-subject conditions, and two additional panels displaying the original results. Each panel has the stimulus TD on the abscissa, with the fastest TD, and most /b/-like stimuli (i.e., step 1), at left, and the slowest TD, and most /w/-like stimuli (i.e., step 11), at right. Stimulus ART is on the ordinate, with the fastest, most /b/-like ART at the top and the slowest, most /w/-like ART at the bottom. The top left corner is thus the most /b/-like stimulus and the bottom right is the most /w/-like, acoustically. The top row thus represents the /b/-amplitude continuum, where ART is held constant at its fastest, and the bottom row represents the /w/-amplitude

continuum, where ART is held constant at its slowest. The diagonal represents the standard continuum, where TD and ART covary. The proportion of responses to the stimulus in the relevant bin is represented by the intensity of the color of the square, with the darkest color representing 100% of that response type, and the lightest (pure white) representing 0%. Note that stops are in red, fricatives in yellow and approximants in blue. Additionally, a combined heatmap, which displays the combination of these three major manner responses via the mixing of these colors, is also shown (Fig. 8).

For each of the six between-subjects conditions, a Bayesian multinomial logistic regression was fit to the manner class of responses, with approximant responses treated as a baseline. Step (centered and scaled) and continuum acted as the predictors, with a full random effects structure, including slopes and intercepts by participant for both step and continuum.

Visually, it appears that listeners make a kind of tripartite division of the stimulus space, with approximant responses throughout the slow-TD region (i.e., the right half of each panel), stop responses in the fast-ART, fast-TD region (i.e., the top left corner of each panel), and fricative responses in the slow-ART, fast-TD region (i.e., the bottom left corner of each panel). This pattern is borne out statistically.

The vertical division of the stimulus space between approximants and non-approximants is represented by a credibly negative effect of step for both stop (Blocked /a/: β = -3.00, Est.Error = 0.33, CrI: [-3.67, -2.35] ; Blocked /i/: β = -3.47, Est.Error = 0.38, CrI: [-4.19, -2.70]; Partial /a/: β = -2.09, Est.Error = 0.24, CrI: [-2.58, -1.65]; Partial /i/: β = -2.41, Est.Error = 0.27, CrI: [-2.94, -1.89]; Unblocked /a/: β = -2.21, Est.Error = 0.35, CrI: [-2.89, -1.53]; Unblocked /i/ β = -2.36, Est.Error = 0.25, CrI: [-2.89, -1.90]) and fricative (Blocked /a/: β = -3.02, Est.Error = 0.40, CrI: [-3.82, -2.20]; Blocked /i/: β = -2.15, Est.Error = 0.40, CrI: [-2.92, -1.34]; Partial /a/: β = -2.30, Est.Error = 0.28, CrI: [-2.83, -1.74]; Partial /i/: β = -1.48, Est.Error = 0.26, CrI: [-1.96, -0.96]; Unblocked /a/: β = -2.21, Est.Error = 0.41, CrI: [-3.02, -1.42]; Unblocked /i/: β = -1.55, Est.Error = 0.22, CrI: [-1.96, -1.09]) responses, across all between-subjects conditions. The

Figure 5: Heatmap of approximant (“w”, “r”, “l”) responses to all stimuli. Within a panel, the abscissa represents TD and the ordinate ART, with TD becoming longer as one moves right on the graph, and ART becoming longer as one moves down on the graph. Each colored cell, then, represents a stimulus, and each line of colored cells represents a continuum: the top line represents the /b/-amplitude continuum, the bottom line represents the /w/-amplitude continuum, and the slanted line represents the standard continuum. Bluer (darker) cells displayed more approximant responses; whiter (lighter) cells display fewer approximant responses. Panels represent different between-subjects conditions, with the original results included at right for comparison. Note that all conditions show more approximant responses at when ART and TD are both longer (i.e., bottom right of the graph is bluer), but only the replication also shows approximant responses when ART is shorter, so long as TD remains long (i.e., the top right is also bluer).

Figure 6: Heatmap of stop (“b”, “d”, “g”) responses to all stimuli. The structure of the figure is similar to Figure 5, except that redder (not bluer) cells indicate more stop responses to that stimulus. Note the concentration of stop responses at fast-ART, fast-TD regions of the stimulus space (i.e., top left is red) in both the replication and the original, with additional stop responses at the fast-ART, slow-TD region in the original (i.e., top right is also red).

Figure 7: Heatmap of fricative (“v”, “f”, “s”) responses to all stimuli. The structure of the figure is similar to Figures 5 and 6, except that yellower (not redder or bluer) cells indicate more fricative responses to that stimulus. Note the consistent concentration of fricative responses at slow-ART, fast-TD regions of the stimulus space, in the original and in the replication.

Figure 8: A heatmap of responses to all stimuli, overlaying the data displayed in Figures 5, 6 and 7 by mixing the colors used in each. Thus, redder cells indicate more stop responses, yellower cells indicate more fricative responses, and bluer cells indicate more approximant responses. Combinations of these colors indicate mixed responses.

Fig. 5: Heatmap of Approximant Responses

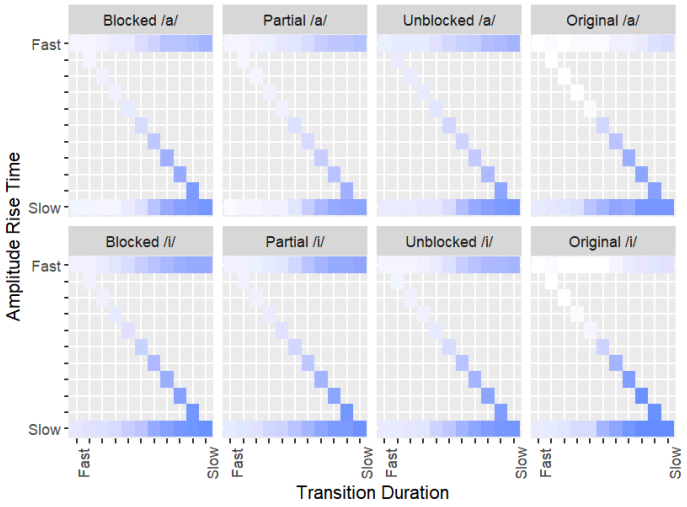


Fig. 6: Heatmap of Stop Responses

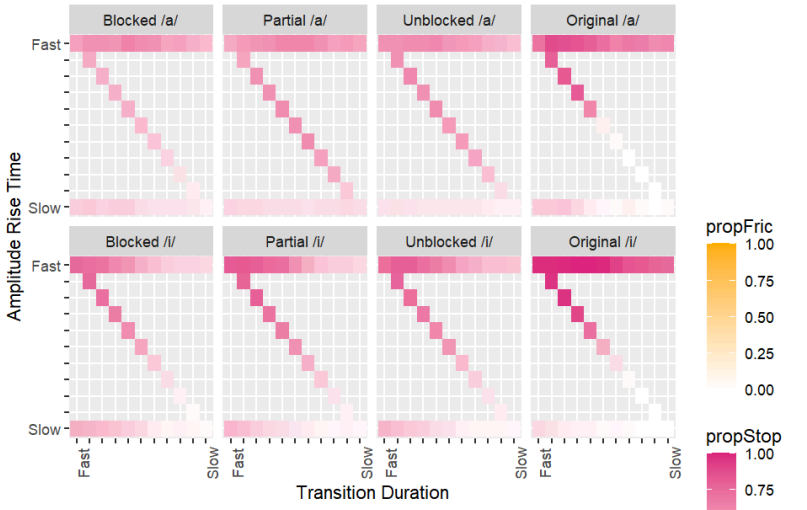


Fig. 7: Heatmap of Fricative Responses

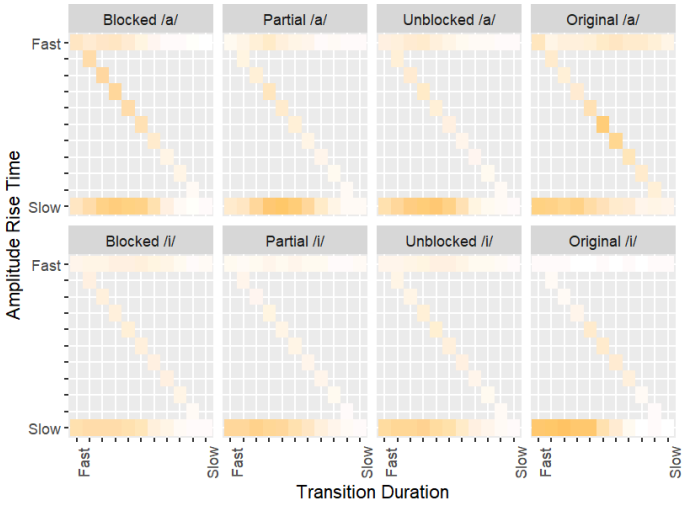
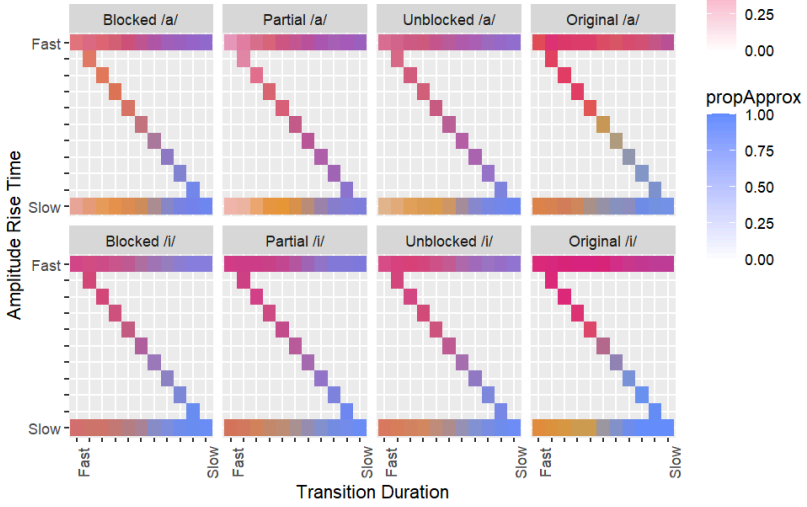
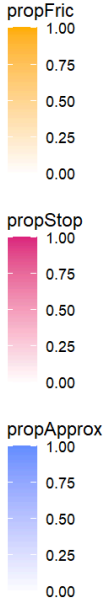


Fig. 8: Free Responses by Manner Class





credibly negative effect implies that, as continuum step (i.e., TD) increases, participants were less likely to say they heard a stop or fricative (and, more likely to say they heard an approximant); in other words, a vertical cut in the stimulus space, dividing approximants, at right, from others, at left.

The horizontal division between stop and fricative responses is generally represented by the models as credible effects of /b/- or /w/-amplitude, of opposite polarity. Stops generally show a credibly negative effect of /w/-amplitude (Blocked /a/: β = -0.92, Est.Error = 0.55, CrI: [-2.06, 0.18]; Blocked /i/*: β = -1.87, Est.Error = 0.41, CrI: [-2.71, -1.10]; Partial /a/*: β = -1.95, Est.Error = 0.43, CrI: [-2.81, -1.13]; Partial /i/*: β = -2.15, Est.Error = 0.33, CrI: [-2.81, -1.54]; Unblocked /a/*: β = -2.17, Est.Error = 0.45, CrI: [-3.09, -1.31]; Unblocked /i/*: β = -2.17, Est.Error = 0.35, CrI: [-2.87, -1.49])[8], which express that stops are less frequent responses in the /w/-amplitude continuum as compared to the standard amplitude continuum. More rarely, stop responses also show a credibly positive effect of /b/-amplitude (Blocked /a/*: β = 1.54, Est.Error = 0.66, CrI: [0.23, 2.88]; Blocked /i/: β = -0.00, Est.Error = 0.30, CrI: [-0.61, 0.57]; Partial /a/: β = 0.20, Est.Error = 0.13, CrI: [-0.03, 0.46]; Partial /i/*: β = 0.21, Est.Error = 0.11, CrI: [-0.01, 0.45]; Unblocked /a/: β = 0.09, Est.Error = 0.13, CrI: [-0.17, 0.37]; Unblocked /i/*: β = 0.60, Est.Error = 0.13, CrI: [0.35, 0.87]), which expresses that stops are more frequent in the /b/-amplitude continuum relative to the standard amplitude continuum. In no case do stop responses show credibly positive effects of /w/-amplitude or credibly negative effects of /b/-amplitude.

Fricative responses, in contrast, generally show credibly positive effects of /w/-amplitude (Blocked /a/: β = 0.11, Est.Error = 0.66, CrI: [-1.23, 1.42]; Blocked /i/: β = 0.51, Est.Error = 0.57, CrI: [-0.59, 1.70]; Partial /a/*: β = 0.96, Est.Error = 0.27, CrI: [0.43, 1.51]; Partial /i/*: β = 0.82, Est.Error = 0.26, CrI: [0.25, 1.31]; Unblocked /a/*: β = 1.13, Est.Error = 0.31, CrI: [0.51, 1.74]; Unblocked /i/*: β = 1.01, Est.Error = 0.33, CrI: [0.37, 1.69]), which express that

[8] I've included the effect of /b/- and /w/-amplitude in all conditions, here, even if not credibly non-zero. Effects which are credibly non-zero are noted with an asterisk.

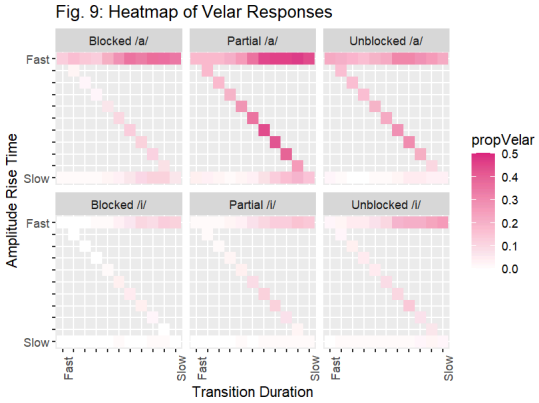
fricatives are more frequent in the /w/-amplitude continuum. Fricative responses also, more rarely, show a credibly negative effect of /b/-amplitude (Blocked /a/: β = -1.06, Est.Error = 0.68, CrI: [-2.51, 0.26]; Blocked /i/: β = -0.33, Est.Error = 0.55, CrI: [-1.42, 0.78]; Partial /a/*: β = -0.44, Est.Error = 0.16, CrI: [-0.76, -0.15]; Partial /i/*: β = -0.43, Est.Error = 0.18, CrI: [-0.80, -0.09]; Unblocked /a/*: β = -0.36, Est.Error = 0.16, CrI: [-0.68, -0.07]; Unblocked /i/: β = -0.04, Est.Error = 0.15, CrI: [-0.35, 0.26]), which express that fricatives are less frequent in the /b/-amplitude continuum. Fricative responses never show a credibly negative effect of /w/-amplitude, nor a credibly positive effect of /b/-amplitude; thus, they demonstrate the reverse pattern of stop responses.

While the specifics differ, most models demonstrate that stop responses are more common in the /b/-amplitude continuum and less common in the /w/-amplitude continuum; or that fricative responses are less common in the /b/-amplitude continuum and more common in the /w/-amplitude continuum. Note that only the Blocked conditions fail to follow this pattern exactly: the Blocked /a/ condition shows only that stop responses are more common in the /b/-amplitude continuum, and the Blocked /i/ condition shows only that they are less common in the /w/-amplitude continuum.

It is worth noting that certain models show credibly non-zero intercepts for either stop or fricative responses. In the Partial /a/ and Unblocked /i/ conditions, the intercept for stop responses is credibly positive (Partial /a/: β = 1.31, Est.Error = 0.41, CrI: [0.53, 2.14]; Unblocked /i/: β = 0.43, Est.Error = 0.18, CrI: [0.07, 0.81]). This implies that stop responses are just slightly more common than approximant responses in these conditions, which is roughly reflected in the intensity of the red (stop) responses being stronger than the intensity of the blue (approximant) responses in the above graphics. In far more conditions, excluding only the Blocked and Partial /a/ conditions, the intercept for fricative responses is credibly negative (Blocked /i/: β = -2.57, Est.Error = 0.97, CrI: [-4.58, -0.78]; Partial /i/: β = -2.04, Est.Error = 0.54, CrI: [-3.14, -0.99]; Unblocked /a/: β = -1.63, Est.Error = 0.78, CrI: [-3.23, -0.16]; Unblocked /i/: β = -1.95, Est.Error = 0.64, CrI: [-3.20, -0.70]). This implies that fricative

responses were generally less common than approximant responses. This latter phenomenon could have a couple underlying causes. It's possible that listeners were less confident in applying fricative labels because the stimuli were not particularly good exemplars of those categories, but this may also arise due to the undercounting of fricative responses. As some participants did not realize non-"b" and non-"w" response alternatives were available, this could also contribute to the lower frequency of fricative responses. It is also possible that, while some stimuli were good exemplars, the stimulus space may transit only briefly through the relevant category.

The interactions found in certain between-subjects conditions can shed some additional light. In all of the /a/ conditions, and two of the /i/ conditions, we find a credibly positive interaction between /b/-amplitude and step for stop responses (Blocked /a/: $\beta = 1.28$, Est.Error = 0.31, CrI: [0.65, 1.89]; Partial /a/: $\beta = 0.79$, Est.Error = 0.17, CrI: [0.47, 1.14]; Unblocked /a/: $\beta = 0.64$, Est.Error = 0.19, CrI: [0.29, 1.01]; Blocked /i/: $\beta = 1.60$, Est.Error = 0.35, CrI: [0.90, 2.27]; Unblocked /i/: $\beta = 0.80$, Est.Error = 0.16, CrI: [0.47, 1.12]). This implies that, relative to the standard continuum, the /b/-amplitude continuum seems to have a flatter stop response curve. In the /a/ conditions, this likely reflects the consistently moderate stop responses throughout the /b/-amplitude continuum, which comes about in part due to the presence of velar responses (Fig. 9) in the fast-ART, slow-TD region of the stimulus space. In the /i/ conditions, this likely reflects the additional stop responses found at the high-TD end of the continuum, which are not found in the corresponding region of the standard continuum, as seen in the redder slow-TD endpoint of the /b/-amplitude continua relative to the standard continuum in Fig. 6, where the fast-TD endpoints are similarly intensely red. In other words, in the /a/ conditions, this probably reflects flattening of the stop response curve at both the high and low end, whereas, in the /i/ conditions, this probably reflects the reduced span caused by only one end of the continuum, the slow-TD end, raising relative to the standard continuum. While, in both cases, stop responses are more common towards the slow-TD endpoint, the origin of this effect may be heterogenous.



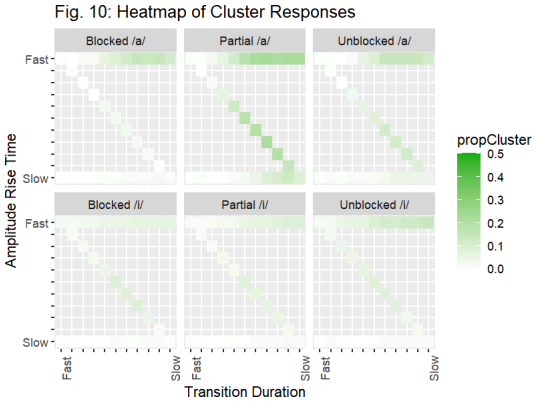


Figure 9: Heatmap of velar responses (“g”, “k”, “gw”) with a similar structure to Figures 5-8, and with redder cells representing more velar responses. Note that the color scale has changed: the maximum proportion is 50%, not 100% as it is in Figures 5-8. Take note of the concentration of velar responses in the fast-ART, slow-TD region of the stimulus space.

Figure 10: Heatmap of cluster responses (“gw”, “kl”, “bw”) with an identical structure to Figures 5-9, and with greener cells representing more cluster responses. Note the color scale has changed, as in Figure 9, to have a maximum of 50%. Note the concentration of cluster responses in the fast-ART, slow-TD region of the stimulus space.

In three conditions, we find a similar interaction for fricative responses, which again seems to show flattening of fricative response curves in the /b/-amplitude continuum. We find a credibly positive interaction of /b/-amplitude and step in the Blocked /i/ (β = 0.98, Est.Error = 0.44, CrI: [0.10, 1.82]), Partial /a/ (β = 0.32, Est.Error = 0.16, CrI: [-0.00, 0.64]) and Unblocked /a/ (β = 0.52, Est.Error = 0.23, CrI: [0.11, 1.00]) condition. Note that, in two cases (Blocked /i/ and Unblocked /a/), this co-occurs with a credibly negative fricative response intercept. Given the relatively few responses in both the standard and /b/-amplitude continuum, and the reduced fricative responses overall in these conditions, this likely just reflects a slightly flatter response distribution, and may not be particularly meaningful.

Finally, a few conditions have scattered interactions of /w/-amplitude and step, where fricative responses generally demonstrate a credibly negative interaction and stops generally demonstrate a credibly positive interaction. In the Blocked /i/ condition alone, stop responses demonstrate a credibly positive interaction of /w/-amplitude and step (β = 1.20, Est.Error = 0.37, CrI: [0.47, 1.96]). This probably reflects the reduced span in this condition triggered by

consistently low stop responses, even at the faster-TD endpoints, as compared with both the standard continuum in the same condition, and the /b/-amplitude continua in other conditions. Note how such responses never become particularly common (i.e., the red in the /w/-amplitude continuum of this condition never becomes particularly intense). We also find credibly negative interactions between /w/-amplitude and step for fricative responses in the Partial /i/ (β = -0.53 , Est.Error = 0.16, CrI: [-0.84, -0.19]) and Unblocked /a/ (β = -0.82, Est.Error = 0.38, CrI: [-1.58, -0.08]) conditions. This likely reflects a small increase in span or sharpness, though this is difficult to detect on visual inspection of the graphs.

To sum up, all conditions clearly demonstrate that stop and fricative responses occur at faster-TD steps of the continua than do approximant responses. Stop responses are generally more common in the /b/-amplitude continuum and less common in the /w/-amplitude continuum, where fricative responses are generally less common in the /b/-amplitude continuum and more common in the /w/-amplitude continuum, though the Blocked conditions only demonstrate differences in the stop responses. This flatter distribution of fricative responses in the Blocked conditions may be a range effect, as fricative responses are clearly more common in the /w/-amplitude continua in all Non-Blocked conditions. Regardless, fricative responses are somewhat less common than approximant responses, perhaps signalling that listeners are uncertain of what category label to apply to these stimuli, though this could also perhaps be due to some participants misunderstanding the free response task. Additionally, stop responses show a reduced span in the /b/-amplitude continua, though this effect seems to be caused by heterogeneous response patterns, with the /a/ conditions demonstrating flatter response patterns across the board, facilitated by the presence of velar responses, and the /i/ conditions demonstrating a higher slow-TD endpoint.

As a final note, the largest difference between the free responses of the original study and the results of our study is in the fast-ART, slow-TD region of the stimulus space. In the original study, they obtain largely stop responses, where in ours we obtain largely approximant responses (compare the top right portions of the panels in Figs. 5, 6, 8). Our

study, thus, demonstrates the largely TD-dependent boundary for the /b/-/w/ distinction, as described above, in line with Nittrouer & Studdert-Kennedy (1986) and Walsh & Diehl (1991), and in clear opposition to the results of the original, which were used to argue for a largely ART-dependent boundary for the /b/-/w/ distinction. This difference between our results and the results of the original study may be due to the change in synthesis parameters, which significantly alter the amplitude and bandwidth values for all five formants. Very early pilot experiments may affirm this possibility. As a consequence, fine details about formant amplitude and bandwidth, or the knock-on effects of altering these synthesis parameters, may be relevant for manner perception.

However, both the original stimulus set and ours elicited cluster responses in that same region of the stimulus space, and nowhere else. These are shown in Fig. 10, above. Note that the total number of cluster responses is relatively low in our results, making up only 4.68% (3312/70758) of total responses, but that they make up a plurality of responses for some stimuli in some conditions (e.g., the eighth step of the /b/-amplitude continuum in the Partial /a/ condition, which has the most cluster responses of any stimulus, has 22.69% cluster responses). This is broadly similar to the results of the original, though there are a few prominent differences worth considering. First, both studies demonstrate a higher rate of cluster responses in some conditions over others, but the distribution of cluster responses is inconsistent between studies. In our study, we see the most cluster responses in the /a/ conditions, particularly the Partial /a/ condition. In the original, it was the /i/ condition that elicited the most cluster responses, though both conditions show a higher proportion of cluster responses than any of our conditions. Further, the majority of responses in that region of the stimulus space consisted of clusters in the original, where it is only a prominent minority in our data. Secondly, there is a difference in the quality of the cluster responses. The original study reports largely labial responses (e.g., “BL”, “BR”, “BW”), whereas we find largely velar responses (e.g., “gl”, “kl”, “qu”). While they do not perfectly overlap, most cluster responses in our data begin with a letter indicating a velar (or, minimally, dorsal) stop.

4 Discussion

4.1 Replication

In replicating Shinn & Blumstein (1984), this study intends to help settle a longstanding debate over the relative importance of ART and TD in the perception of the /b/-/w/ distinction for American English listeners. In the original study, only the standard continuum was found to produce a sigmoidal categorization function, where the /b/-amplitude continuum instead largely returned “b” responses throughout and the /w/-amplitude continuum largely returned “w” responses throughout. This was taken to mean that ART was the more important correlate of the two to American English listeners. Their free response task then demonstrated that the stimulus space elicited unexpected responses, which they believe were largely engendered by the high frequency energy present in the initial formant positions. From this they concluded that ART was still likely an invariant in stop-approximant distinctions.

However, their conclusion is somewhat incongruous with the current data. Were ART the sole (or primary) correlate of the distinction, we should be seeing near-ceiling “b” responses when ART is held constant at a faster rate, and near-floor “b” responses when ART is held constant at a slower rate, and this is simply not found. The results of the categorization task seem to indicate that the standard and /b/-amplitude continua behave similarly, at least when blocked together, and are thus likely categorized on the basis of a common criterion, probably related to TD. While the flattening of the /w/-amplitude continuum across conditions does look like the original, this does not appear in the Partial /a/ and Unblocked /a/ conditions, and the depression may be related to the presence of a third category, so /w/-amplitude flattening may not derive from the use of an ART-based criterion. Overall, regardless of interpretation, our results, particularly the results from the /b/-amplitude continuum, are quite dissimilar from those of the original study, even in the Blocked condition which most closely resembles the blocking structure of the original study.

Additionally, the results of the free response task seem to indicate that the interaction of ART and TD is more complicated than the original paper suggests. While longer TDs are associated with “w” responses regardless of ART, which would suggest that ART is less relevant in their categorization than TD, shorter TDs are only strongly associated with “b” responses when ART is also short. Otherwise, “v”-like responses tend to dominate. This pattern of results was also found in the original study, less the approximant responses in the fast-ART, slow-TD region, but “v” responses were argued to be triggered by differences in initial gross spectral character, as demonstrated for categorization of stop place in Lahiri et al. (1984), and thus distinct from (and incidentally confounded with) the correlates discussed. Note that, in replicating their stimuli, we cannot disconfirm the hypothesis that the gross spectral character of the initial spectra provided evidence of a non-labial place of articulation, though the fact that changes in formant amplitude and bandwidth settings imposed by the cascade branch of the Klatt synthesizer and by experimenter error, respectively, did not alter the presence of fricative responses may hint that this finding could generalize to other gross spectral shapes.

However, given the present categorization responses fail to reflect invariant ART as discussed in Shinn & Blumstein (1984), but the free response results are highly similar to those obtained by the original study, again, less the fast-ART, slow-TD region, it seems ART and TD may interact in more complex ways than previously described. TD seems to be more important in the categorization of approximant-like “w” responses, regardless of ART, but ART is still relevant in the distinction of /b/ and /v/. (Notice, this is also reflected in the categorization results: participants consistently rate the slower-TD end of all continua across all conditions as “w”, but are more uncertain about the faster-TD end of the continua.) While the notion that ART forms an invariant cue distinguishing stops from continuants seems unlikely, the possibility that ART and TD provide information about consonant manner certainly seems accurate.

4.2 Fricative Perception

Perhaps the most striking result of this and the original experiment is the presence of fricative responses across all between-subjects conditions to the slow-ART, fast-TD region of the stimulus space. None of the stimuli contain aperiodic energy, which is traditionally understood as a primary correlate for fricatives. Indeed, if the current data are to be believed, it seems sufficient for fricative categorization for a sound to have the right profile of periodic information alone.

This is, to a degree, in line with Harris's (1958) proposal, which argues that the non-sibilant fricatives are evinced largely via periodic information, but would require that non-sibilant fricative categories be distinguished from non-fricatives without the use of frication (or, would require that non-sibilant fricatives be classed entirely separately from sibilant fricatives). While frication intensity has been argued to cue sibilancy (Harris, 1958, McMurray & Jongman, 2011), it is unclear how "fricativeness" could be signaled by non-sibilants. There are two central possibilities: either non-sibilant fricatives are straightforwardly evinced by differences in temporal dynamics (i.e., amplitude profile or formant profile, as manipulated above, is itself taken as evidence for categorizing sounds as fricatives, even when non-sibilant frication is unavailable) or non-sibilant fricative categories are selected by listeners due to other calculations (e.g., of place), which limits the alternatives available to a listener (e.g., via the gross spectral prominence (Lahiri et al., 1984), or via the intensity and distribution of apparent noise (McMurray & Jongman, 2011)). It is also possible that some element of the stimuli is causing a percept of aperiodicity or "buzzing," which is then mistaken for frication. Regardless, so long as listeners are indeed arriving at a non-sibilant fricative category, there must be some purely periodic information which signals that a non-sibilant fricative is present.

That said, there is a reasonable objection which one might raise about these results: perhaps listeners are not so much convinced that they are hearing a /v/ as convinced they are not hearing a /b/ or /w/. Perhaps listeners are not hearing a good exemplar of /v/ but,

without a more appropriate category, and clearly hearing something which does not sound like a good exemplar of /b/ or /w/, listeners land on the most appropriate (i.e., least wrong) category: /v/. The current data cannot directly legislate on this, as uncertainty cannot be simply or rigorously verified, but participants familiar with Hindi-Urdu and related languages remarked that the /v/-like sounds were similar to /ʊ/. Future work will need to examine this and related stimulus spaces for not only the category labels listeners are willing to provide, but for their confidence in providing those labels.

If, indeed, these are good exemplars of a fricative category, this would also contradict the position that fricatives should not undergo diachronic dissimilation because they are evinced by different information than that which evinces vowels. Ohala (1981) argues that listeners cause dissimilation via over-correction of coarticulatory noise, which erroneously attributes the featural information of one segment to another. We could interpret this to mean that only adjacent segments which share acoustic information should trigger dissimilation. Because of the then-common idea that "none of [the features [fricative], [affricate] and [stop]] have major perceptual cues on adjacent segments" (Ohala, 1981:194), it followed that listeners would never correct for coarticulatory noise between the localized frication characterizing fricatives and the periodic information of adjacent vowels. Following this reasoning, one might predict that fricatives are unlikely to dissimilate. With no shared acoustic information between fricatives and vowels –or, at least, none a listener was believed to use– fricatives should be exempt from diachronic dissimilation. The above results, if they hold, upset the position that fricatives and vowels share no key acoustic information: it seems clear that variations in ART and TD can cause the appearance and disappearance of fricative judgments without categorically adding or removing acoustic information. Thus, we must either abandon the idea that non-sibilant fricatives are fricatives, or retract the claim that fricatives will never trigger diachronic dissimilation.

Even more broadly, these results suggest a tripartite division in ART-TD space may be used by American English listeners in categorizing bilabial consonants' manner, which may

suggest that such a division could exist in some capacity for the perception of consonant manner across consonant place, in some language. However, as shown by Benkí (2001), consonant place and voicing can be perceptually correlated, so there may be non-identity (or nonlinear) relationships between manner and place categories, and thus we have no guarantee that this tripartite division will be found for manner contrasts in different places within the same language, much less found consistently across languages.

5 Conclusion

While the conclusion of the original paper, that ART acts as an invariant cue for the /b/-/w/ distinction in American English, does not seem to be reflected in the results of the present study, the core idea that ART and TD have a strong effect on manner perception seems to be validated. If we abandon the notion that speech perception relies on invariant speech codes, and instead adopt the perspective that multiple sources of information are used to make categorization decisions, it becomes clear that the free response results of both the present and original studies point to the use of ART and TD –the rate of change of intensity and formant center frequency slope– as a core component in our decision-making about consonant category. This opens up the possibility that information about the trajectory of formant information is perhaps used widely, as has already been suggested in work on silent center effects and theories of vowel inherent spectral change (Jenkins et al., 1983; Morrison, 2012; Nearey & Assmann, 1986; Nearey, 2012; Strange et al., 1983; Strange & Jenkins, 2012).

If it is true that formant trajectory is, indeed, used in the categorization of consonants and vowels, and that the categorization of even fricatives relies on formant trajectory, there then comes a question of precisely what information about formant trajectory is stored by listeners of the world's languages. Some suggest that the entire trajectory could be helpful to some degree in vowel classification (Watson & Harrington, 1999; Zahorian & Jagharghi, 1991), and may be stored in some capacity for use in categorization (Zahorian & Jagharghi, 1993), while

others suggest only a small number of guiding points are sufficient to effectively model vowel category prototypes (Morrison, 2012). We could also imagine a Motor Theoretic analysis-by-synthesis model (Liberman & Mattingly, 1985) or a record not of the auditory trajectory itself but of the parameters required to recapitulate it. Further, it is as of yet unclear if formant trajectory information could interact (or, trade) with other known consonant correlates, particularly the presence, absence and intensity of aperiodic energy as would potentially signal a fricative. This may, then, be a site for fruitful future investigation.

Lastly, and most speculatively, it appears that the notion of a fricative is not internally perceptually consistent –and may, in fact, be perceptually ill-defined. Almost no work has directly questioned how fricatives are diagnosed as such, leaving a massive gap in our knowledge, but the results above tentatively suggest that non-sibilant fricatives, or highly related categories, can be diagnosed entirely from periodic information, without invoking their characteristic aperiodicity. If this crucial element of the signal can simply be excised, then perhaps fricative and approximant categories are far more fluidly related than previously thought.

References

Barreda, S. (2016). Investigating the use of formant frequencies in listener judgments of speaker size. *Journal of Phonetics*, *55*, 1-18. doi.org/10.1016/j.wocn.2015.11.004.

Baumann, O. & Belin, P. (2008). Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychological Research PRPF*, *74*, 110-120. doi.org/10.1007/s00426-008-0185-z.

Behrens, S., & Blumstein, S. E. (1988). On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricatives. *Journal of the Acoustical Society of America*, *84*(3), 861-867.

Benkí, J. R. (2001). Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics*, *29*(1), 1-22. doi.org/10.1006/jpho.2000.0128.

Boersma, P. & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.08, retrieved 10 February 2024 from http://www.praat.org/.

Brady, S. A., & Darwin, C. J. (1978). Range effect in the perception of voicing. *Journal of the Acoustical Society of America*, *63*, 1556-1558. doi.org/10.1121/1.381849.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M. & Gerstman, L. J. (1952). Some Experiments on the Perception of Synthetic Speech Sounds. *Journal of the Acoustical Society of America*, *24*, 597-606. doi.org/10.1121/1.1906940.

Diehl, R. L., Elman, J. L. & McCusker, S. B. (1978). Contrast Effects on Stop Consonant Identification. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 599-609./doi/10.1037/0096-1523.4.4.599.

Eimas, P. D. (1963). The Relation between Identification and Discrimination along Speech and Non-Speech Continua. *Language and Speech*, *6*(4), 206-217. doi.org/10.1177/002383096300600403.

Fant, G. (1960). *Acoustic Theory of Speech Production: With Calculations based on X-Ray Studies of Russian Articulations*. Mouton, The Hague.

Harris, K.S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, *1*, 1-7.

Hedrick, M. S., & Ohde, R. N. (1993). Effect of relative amplitude of frication on perception of place of articulation. *Journal of the Acoustical Society of America*, *94*, 2005-2026. doi.org/10.1121/1.407503.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, *72*(5), 1218-1227. doi.org/10.3758/APP.72.5.1218.

Jenkins, J. J., Strange, W. & Edman, T. R. (1983). Identification of vowels in “vowelless” syllables. *Perception & Psychophysics*, *34*, 441-450. doi.org/10.3758/BF03203059.

Kent, R. D. & Vorperian, H. K. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders*, *74*, 74-97. doi.org/10.1016/j.jcomdis.2018.05.004.

Kingston, J. & Rysling, A. (2023). When is Enough, Enough? VOT Judgments Vary by Voicing Intensity, Aspiration Intensity, Voice Quality, and Rate of Change. *Proceedings of the International Congress of Phonetic Sciences*, Prague, Czech Republic. www.dropbox.com/s/u3c5xvxejbpb2ag/paper1080_KingstonRysling_Final_28apr23.pdf?dl=0

Kingston, J., & Rysling, A. Perceptual effects of context intensity [Unpublished manuscript].

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*(3), 971-995. doi.org/10.1121/1.383940.

Kluender, K. R., & Walsh, M. A. (1992). Amplitude rise time and the perception of the 'voiceless affricate/fricative distinction. *Perception & Psychophysics*, *51*(4), 328-333. https://doi.org/10.3758/BF03211626.

Lahiri, A., Gewirth, L., & Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants. *Journal of the Acoustical Society of America*, *76*, 391-404. doi.org/10.1121/1.391580.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the Speech Code. *Psychological Review*, *74*(6), 431-461. doi/10.1037/h0020279.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1-36. doi.org/10.1016/0010-0277(85)90021-6.

Mack, M. & Blumstein, S. E. (1983). Further evidence of acoustic invariance in speech production: The stop-glide contrast. *Journal of the Acoustical Society of America*, *73*, 1739-1750. doi.org/10.1121/1.389398.

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*, 219-246

Miller, G. A. & Nicely, P. E. (1954). An Analysis of Perceptual Confusions among English Consonants Heard in the Presence of Random Noise. *Journal of the Acoustical Society of America*, *27*, 953. doi.org/10.1121/1.1928048.

Miller, G. A. & Nicely, P. E. (1955). An Analysis of Perceptual Confusions Among Some English Consonants. *Journal of the Acoustical Society of America*, *27*, 338-352. doi.org/10.1121/1.1907526.

Miller, R. L., (1953). Auditory Tests with Synthetic Vowels. *Journal of the Acoustical Society of America*, *25*, 114-121. doi.org/10.1121/1.1906983.

Miller Willahan, C. (2023). *How Unexpected: Exploring the Effect of Phonological Features on Perception of Sound Errors* [Master's thesis, UC Santa Cruz]. ProQuest Dissertations Publishing. Retrieved from escholarship.org/uc/item/2t7933rw.

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing* (5th ed.). Elsevier Science.

Morrison, G. S. (2012). Theories of Vowel Inherent Spectral Change. In G. Morrison, P. Assmann (eds.), *Vowel Inherent Spectral Change* (pp. 31-47). Springer. doi.org/10.1007/978-3-642-14209-3_3.

Nearey, T. M. & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, *80*, 1297-1308. doi.org/10.1121/1.394433.

Nearey, T. M. (2012). Vowel Inherent Spectral Change in the Vowels of North American English. In G. Morrison, P. Assmann (eds.), *Vowel Inherent Spectral Change* (pp. 49-85). Springer. doi.org/10.1007/978-3-642-14209-3_4.

Nittrouer, S., & Studdert-Kennedy, M. (1986). The Stop-Glide Distinction: Acoustic Analysis and Perceptual Effect of Variation on Syllable Amplitude Envelope for Initial /b/ and /w/. *Journal of the Acoustical Society of America*, *80*, 1026-1029. doi.org/10.1121/1.393843.

Nittrouer, S., & Studdert-Kennedy, M. (1987). The Role of Coarticulatory Effects in the Perception of Fricatives by Children and Adults. *Journal of Speech, Language, and Hearing Research*, *30*(3), 319-329. doi.org/10.1044/jshr.3003.319.

Nittrouer, S. (2002). Learning to perceive speech: How fricative perception changes, and how it stays the same. *Journal of the Acoustical Society of America*, *112*, 711-719. doi.org/10.1121/1.1496082.

Ohala, J. J. (1981). The Listener as a Source of Sound Change. *Parasession on Language and Behavior/Chicago Linguistic Society*. linguistics.berkeley.edu/~ohala/papers/listener_as_source.pdf.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203. dx.doi.org/10.3758/s13428-018-01193-y.

Peterson, G. E. & Barney, H. L. (1952). Control Methods Used in a Study of the Vowels. *Journal of the Acoustical Society of America*, *24*, 175-184. doi.org/10.1121/1.1906875.

R Core Team (2023). R [Computer Program].

Repp, B. H. (1979). Relative Amplitude of Aspiration Noise as a Voicing Cue for Syllable-Initial Stop Consonants. *Language and Speech*, *22*(2), 173-189. doi.org/10.1177/002383097902200207.

Rosen, S. M. (1979). Range and frequency effects in consonant categorization. *Journal of Phonetics*, *7*, 393-402. doi.org/10.1016/S0095-4470(19)31072-1.

Schorer, E. (1986). Critical modulation frequency based on detection of AM versus FM tones. *Journal of the Acoustical Society of America*, *79*, 1054-1057. doi.org/10.1121/1.393377.

Shinn, P. & Blumstein, S. E. (1984). On the Role of the Amplitude Envelope for the Perception of [b] and [w]. *Journal of the Acoustical Society of America*, *75*(4), 1243-1252. doi.org/10.1121/1.390677.

Stevens, K. N. (1985). Evidence for the role of acoustic boundaries in the perception of speech sounds. In V. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*. Academic Press.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*(1-2), 3-45. doi.org/10.1016/S0095-4470(19)31520-7.

Strange, W., Jenkins, J. J. & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*, *74*(3), 695-705. doi.org/10.1121/1.389855.

Strange, W., Jenkins, J.J. (2012). Dynamic Specification of Coarticulated Vowels. In G. Morrison, P. Assmann (eds.), *Vowel Inherent Spectral Change* (pp. 87-115). Springer. doi.org/10.1007/978-3-642-14209-3_5.

Sussman, H. M., Bessell, N., Dalston, E., & Majors, T. (1997). An investigation of stop place of articulation as a function of syllable position: A locus equation perspective. *Journal of the Acoustical Society of America*, *101*, 2826-2838. doi.org/10.1121/1.418567.

Sussman, H. M., McCaffrey, H. A. & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America*, *89*(4), 1998. doi.org/10.1121/1.401923.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021) Rank-Normalization, Folding, and Localization: An Improved R for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, *16*(2), 667–718.

Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. *Psychonomic Bulletin & Review*, *16*(1), 74-79. doi.org/10.3758/PBR.16.1.74.

Wagner, A., Ernestus, M. & Culter, A. (2006). Formant transitions in fricative identification: The role of native fricative inventory. *Journal of the Acoustical Society of America*, *120*, 2267-2277. doi.org/10.1121/1.2335422.

Walsh, M. A. & Diehl, R. L. (1991). Formant Transition Duration and Amplitude Rise Time as Cues to the Stop/Glide Distinction. *The Quarterly Journal of Experimental Psychology*, *43*(3), 603-620. doi.org/10.1080/14640749108400989.

Watson, C. I. & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America*, *106*, 458-468. doi.org/10.1121/1.427069.

Zahorian, S. A. & Jagharghi, A. J. (1991). Speaker normalization of static and dynamic vowel spectral features. *Journal of the Acoustical Society of America*, *90*, 67-75. doi.org/10.1121/1.402350.

Zahorian, S. A. & Jagharghi, A. J. (1993). Spectral-shape features versus formants as acoustic correlates for vowels. *Journal of the Acoustical Society of America*, *94*, 1966-1982. doi.org/10.1121/1.407520.

Zwicker, E., & Fastl, H. (1999). *Psychoacoustics: Facts and models* (Vol. 22). Springer Science & Business Media.