Abstract

In four experiments we explored how biphone probability and lexical neighborhood density influence listeners' categorization of vowels embedded in nonword sequences. We found independent effects of each. Listeners shifted categorization of a phonetic continuum to create a higher probability sequence, even when neighborhood density was controlled. Similarly, listeners shifted categorization to create a non-word from a denser neighborhood, even when biphone probability was controlled. Next, using a visual world eye-tracking task, we determined that biphone probability information is used rapidly by listeners in perception. In contrast, task complexity and irrelevant variability in the stimuli interfere with neighborhood density effects. These results support a model in which both biphone probability and neighborhood density independently affect word recognition, but only biphone probability information is encoded prelexically.

1. *Introduction*

Listeners rely on knowledge about the phonological and lexical organization of their language when they process speech. Two such influences are biphone probability - the probability of two sounds occurring in sequence, and lexical neighborhood density - the number and frequency of similar sounding words in the lexicon (Vitevich & Luce, 1999). Both have been shown to influence phonetic categorization. In a series of experiments, we evaluated the independent contribution and time course of biphone probability and neighborhood density effects in a phonetic categorization paradigm.

Lexical neighborhood density is defined as the number of known words that are similar to a string by a given metric. Commonly, a neighbor is defined in terms of phoneme overlap: a word's (or non-word's) neighbors are words which can be created from substituting, adding or deleting a single phoneme. As captured in the Neighborhood Activation Model (Luce and Pisoni 1998), the central idea is that in speech processing, multiple lexical candidates are activated based on their similarity to the input, with various consequences for the processing of both words and non-words. As a result of competition, the recognition of sequences from high density neighborhoods is slower compared to sequences from low density neighborhoods (e.g., Luce & Pisoni, 1999; Vitevitch, 2002a), although the production and recall of sequences from high density neighborhoods is privileged in contrast to those from lower density neighborhoods (e.g., Vitevitch, 2002b; Roodenrys & Hinton, 2002). Children as well recognize words from high density neighborhoods more slowly than those from low density neighborhoods (e.g., Garlock, Walley, & Metsala, 2001; Munson, Swenson & Manthei, 2005). Sensitivity to neighborhood density emerges gradually only during the second year of life. Thus, 14-month-olds are sensitive to the details of pronunciation of familiar words from high as well as low density neighborhoods (Swingley & Aslin, 2002), but by 17-months infants are more likely to learn novel words from low density neighborhoods compared to those from high density neighborhoods (Hollich, Jusczyk & Luce, 2002).

It is typically challenging to distinguish effects of neighborhood density from those of biphone probability because these measures are highly correlated, at least in English (Pitt & McQueen 1998, Vitevitch & Luce 1998; Vitevitch, Luce, Pisoni & Auer, 1999; Landauer & Streeter 1973): words in denser lexical neighborhoods tend to be comprised of higher probability sequences. Nonetheless, whether neighborhood density and biphone probability independently affect speech perception is central to the distinction between theories that do and do not advocate for the role of feedback in models of spoken word recognition.

In this paper we focused on isolating the role of biphone probability and neighborhood density using a phonetic categorization task. To do so, we built on an experiment by Newman et al. (1997). Newman et al. tested phonetic processing using a 2AFC task in which listeners categorized a VOT continua, with two non-word endpoints. They found that listeners categorization of the VOT continua was biased towards non-words from denser neighborhoods. Newman et al., argue that their results can only be captured by models of speech processing which are interactive, and allow for feedback from lexical to prelexical levels of representation. TRACE (McClelland & Elman 1986) represents one such interactive model which implements feedback from activated lexical entries to a lower, phonemic layer of representation. In TRACE, an item with an ambiguous stop consonant activates both non-word endpoints, which in turn activate lexical neighbors. Top-down activation from these neighbors then boosts activation for the denser-neighborhood non-word to a greater extent, biasing categorization in its direction. Thus, Newman et al., argue that their results are supportive of such a model where activation of neighbors modulates the activation of prelexical nodes via feedback.

Norris, McQueen & Cutler (2000) argue that Newman et al.'s results can be explained without recourse to feedback. One possibility they suggest is that Newman et al.'s neighborhood density effects may be attributed to differences in biphone probability alone. It has been previously shown that listeners tend to categorize an ambiguous stop consonant as one that results in a higher probability sequence given the preceding segment (Pitt & McQueen, 1998). Crucially, if such differences can be explained by differences in biphone probability alone, this obviates the need for feedback from the lexicon.

However, Norris et al's hypothesis has been only partially supported. As Newman et al., argue, their results cannot be explained by differences in the probabilities between the initial consonant and the following vowel because they controlled for it. Similarly, Brancazio & Fowler (2000) argue that at least for some continua tested by Newman et al., the neighborhood effects cannot be explained by differences in the probabilities of the non-adjacent consonants; although Norris et al. provide some evidence that Newman et al.'s results could be attributed to higher order (triphone) probabilities.

Alternately, Norris et al., (2000) argue that Newman et al.'s results could be lexical and influence categorization, but at the later decision stage, and thus be modeled as a response bias. Because lexical effects at the decision stage do not alter the early recognition of non-words, feedback is not necessary to explain them. Consistent with this hypothesis, Newman et al. report neighborhood density effects only at intermediate and long (cf. Fox, 1984) but not short reaction times. These late effects of neighborhood density could well emerge from the influence of lexical variables at the decision stage.

Based on these findings, Pitt & McQueen (1998) argue for autonomous models of speech processing where listeners' expectations about sound sequences, as indexed by biphone probabilities, alone feed-forward activate phonemic units, as in Shortlist A (Norris 1994). Further, effects of lexical neighborhoods do not provide interactive feedback from lexical to prelexical levels of processing, but instead influence decisions due to feed-forward activation of decision nodes.

Biphone probability effects are also well established in the literature. Adults are more likely to recognize, name (e.g., Frisch, Large & Pisoni, 2000; Vitevitch, Armbuster & Chu, 2004), recall (Thorn & Frankish, 2005) and accept as word-like (Pierrehumbert, Needle & Hay, 2018), high probability sequences, compared to sequences with a lower probability. This advantage for high probability sequences is evident in children as well who produce nonwords with high probability sequences more accurately (e.g., Munson, Edwards & Beckman, 2005; Gathercole, Frankish, Pickering & Peaker, 1999).

Finally, these effects seem to be evident even in infancy. Whether infants are learning English (Jusczyk, Luce & Charles-Luce, 1994; Mattys, Jusczyk, Luce & Morgan, 1999), Dutch (Freiderici & Wessels, 1993) or Catalan (Sebastian-Gallés & Bosch, 2002), 9-months listen longer to high probability sequences compared to those with a low probability. Additionally, English learning 9-month-olds can use dips in biphone probability sequences to segment words (Mattys & Jusczyk, 2001); they can also segment nonce words beginning with high biphone probability sequences but not those with low biphone probabilities (Archer & Curtin, 2016). In sum, biphone probability effects on speech perception and production are evident early in acquisition and through adulthood.

These two sets of findings thus offer contrasting views of the variables implicated in phonetic processing. In the view advocated by Newman et al (1997), neighborhood activations play a central role in phonetic processing. As exemplified in TRACE, Newman et al attribute these neighborhood effects to feedback from the lexicon. Critically in TRACE, feedback alters the prelexical activation of phones; but there is no independent representation of biphone information. Given the high correlation between biphone probability and neighborhood density, phonotactic probability effects in phonetic processing in such models are simply a by-product of neighborhood activations. That is, a higher density neighborhood increases activation for high probability words and non-words. Further, because feedback introduces a delay, neighborhood density effects are not immediate. However, this account fails to capture how biphone sensitivity in young infants might correspond to neighborhood effects seen in the second year of life.

Alternatively, there are models where it is biphone probabilities that play a central role in phonetic processing. As exemplified in Shortlist A (Norris, 1994) and Merge (Norris, 1999), such models include architecture compatible with a prelexical, autonomous representation of sequential/phonotactic information with no independent role for neighborhood density. Because biphone probability effects are represented prelexically, they are expected to influence phonetic processing with little to no delay.

Finally, Norris et al (2000) outline a third possibility where both biphone probability and neighborhood density independently influence phonetic processing. In this proposal as well, biphone probability influences are prelexical. In addition, neighbors are activated and feed-forward activation to decision nodes. Thus, unlike biphone probability, neighborhood density does not affect the prelexical activation of phones. Instead, it acts as a bias at the decision stage. In this account, neighborhood density effects are delayed relative to biphone probability effects, though not as late as might be expected from a feedback account.

As is clear from the preceding discussion, answers to two questions are critical in teasing apart these accounts. First, are biphone probability and neighborhood density effects independent? Second, what is the time course for biphone probability and neighborhood density effects? In four experiments, we used phonetic categorization of non-words to disentangle the contribution of biphone probability and neighborhood density. Following Newman et al. (1997) and Pitt and McQueen (1998) we used non-words because this allowed us to test listeners' use of information which does not directly depend on word-hood, word frequency, semantic associations with words, and so on. First, we tested whether biphone probability and neighborhood density independently influence phonetic categorization when the other variable is controlled. Next, we used eye-tracking to determine the time course of each of these effects. Together, these results address how and when listeners use lexical and phonological information in speech processing, and thus inform models of spoken word recognition.

2. *Experiment 1: Biphone probability effects controlling for neighborhood density*

The goal of Experiment 1 was to test whether differences in biphone probability influenced listeners' categorization of a continuum, when neighborhood density was controlled. To this end, we created a continuum from the English vowels /ɛ/ to /æ/ by manipulating F1, F2 and F3. The continuum was presented in one of two CVC frames and listeners were asked to categorize the vowel as /ɛ/ to /æ/. The two frames were: /mɛb/~/mæb/ and /mɛv/~/mæv/. The consonant frames were selected such that neither endpoint was a word in English, and both coda consonants /b/ and /v/ involved labial constrictions so formant trajectories at the offset of the vowel could be expected to be similar, allowing for identical continuum steps to be used with each frame. Critically, continuum endpoints were selected to control for neighborhood density biases, while varying BP biases. We first describe how ND and BP metrics were computed in the paper as a whole and then turn to these measures as relevant to Experiment 1.

2.1 Calculation of biphone frequency and neighborhood density

All density and biphone probability measurements were made using the KU Phonotactic Probability Calculator and KU Neighborhood Density Calculator (Vitevich & Luce 2004), which provides frequency-weighted positional estimates for individual phones in a sequence, as well as biphone co-occurrence probabilities. The lexicon used in the calculators is based on the Merriam Webster Pocket Dictionary, with frequency measures from Kučuera & Francis (1967). Neighborhood density was calculated using the same formula as in Newman et al. (1997), where each neighbor's contribution was frequency weighted. Each neighbor's frequency contribution was calculated by taking the logarithm (base 10) of the raw frequency times 10. This value was then summed for all neighbors for a given word, to provide a frequency-weighted neighborhood density. To ensure that the words entered into the calculation were likely known by our participants, we used only words that have previously been rated as familiar (Nusbaum, Pisoni & Davis 1984), using a familiarity index of 5.0 or higher as a cut-off (on a 7-point scale, see Nusbaum et al. 1984). We also made the same calculations including all (even less familiar) words, this did not change the direction of any predicted effects.

We used a second metric to provide an alternative computation for BP. To this end we used a python scripted program that produces positional BP metrics given an input lexicon (Breiss, 2021). We computed these measures using the Carnegie Mellon University Pronouncing Dictionary corpus (Weide 1998), employing the version of the dictionary described and used in Hayes (2012), which includes words with frequencies of at least 1 in the CELEX database (Baayen, Piepenbrock & Gulikers, 1995). The Breiss calculator operates by taking a corpus of text as a transcription system where there is a one-to-one mapping between phone and character, and then creates a dictionary of frequencies for each character, character pair, and character triplet in the corpus. Then for any given sequence, the program looks up the unigram, bigram, and trigram probability for that sequence, multiplying together the subsequences to yield the product of the whole. The code for the program is available on the repository for this paper at: https://osf.io/eba2v/. The KU Phonotactic probability calculator and the Breiss phonotactic probability calculator were in agreement in terms of the directionality of bias differences across consonant frames, with one minor exception in Experiment 2, described below. We take the alignment of the measures as converging evidence for the predicted frame effects in terms of BP.

Table 1 shows the biphone-probabilities and frequency-weighted neighborhood densities for the endpoints of all the continua used in this study. This includes BP information for both C_1V_2 and V_2C_3 in the CVC sequence. It also includes ND for C_1V_2 as well as for the CVC sequence as a whole.

TABLE 1 HERE

2.2 BP and ND metrics in Experiment 1

First consider neighborhood density for the full CVC sequence of the two continua used in Experiment 1: the non-word /mɛb/ has a frequency-weighted neighborhood density of 17.96. The other endpoint of the continuum, /mæb/ has a frequency-weighted neighborhood density of 29.54. Following Newman et al., listeners should be biased towards a denser-neighborhood non-word when exposed to an ambiguous stimulus. In the present case, a denser neighborhood for /mæb/ would bias listeners to respond /æ/ when exposed to ambiguous items on a /mɛb/~/mæb/ continuum. We can index the magnitude of this bias by subtracting the frequency-weighted neighborhood density of /mæb/ from that of /mɛb/. The /mɛb/~/mæb/ continuum therefore has a neighborhood density bias that is negative, i.e., biased towards /æ/. The bias is -11.94 (17.96-29.54). The /mɛv/~/mæv/ continuum also has a neighborhood density bias for /æ/ of (-12.88). Comparing the biases for the two continua, we see that although both have an /æ/ bias, the /mɛv/~/mæv/ continuum has a slightly larger one. This would predict that if listeners are sensitive to neighborhood density alone, they should show increased /æ/ responses to the /mɛv/~/mæv/ continuum compared to /mɛb/~/mæb/ continuum. However it should be noted that the difference in bias across continua here is much smaller than reported for the continua used by Newman et al. (1997), suggesting its influence may be minimal.

Continuum biases were calculated in the same way for biphone probability using the KU Phonotactic probability calculator, for both C_1V_2 and V_2C_3 in the CVC string, though note that the onset consonant is the same across continua in Experiment 1 so it does not contribute to any sequential probability differences across conditions. As shown in Table 1, the V_2C_3 portion of the /mɛb/~/mæb/ continuum exhibited an /æ/ bias (-0.0019), while the V_2C_3 portion of the /mɛv/~/mæv/ continuum exhibited an /ɛ/ bias (0.0007). This differential predicts that a coda /b/ should bias listeners towards /æ/ responses, such that they prefer a relatively higher probability sequence /mæb/ (as compared to /mɛb/), and vice versa for coda /v/. We can note that the metrics computed using the Breiss Calculator for the full CVC sequence are in agreement with this conclusion, where a coda /v/ also biases listeners towards /ɛ/ in Experiment 1.

If listeners are sensitive to biphone probability information, they should thus show *increased* /ɛ/ responses for the /mɛv/~/mæv/ continuum compared to the /mɛb/~/mæb/ continuum, with coda /v/ biasing towards /ɛ/. Note that the bias based on biphone probability is in the opposite direction on the bias predicted by neighborhood density, making this a fairly conservative test for biphone probability effects (though density biases are minimally different). A finding in the predicted direction would therefore provide strong evidence that biphone probability exerts an independent influence on listeners' categorization of speech sounds.

2.3 Materials

Stimuli for Experiment 1 were created by resynthesizing the speech of an adult male speaker of American English. The stimuli were first recorded at 44.1 kHz (32 bit) in a sound-attenuated room, using an SM10A Shure™ microphone and headset. The starting point for the creation of stimuli was the speaker's natural production of two CVC nonwords: /mɛv/ and /mæv/. The vocalic portion of both of these nonwords was excised from the CVC frame. The continuum was then synthesized in Praat via LPC decomposition and resynthesis of F1, F2 and F3 (Winn, 2016). Resynthesis used /ɛ/ as a base and interpolated F1, F2, and F3 in evenly Bark-spaced steps to their respective values for the /æ/ token in 12 steps. The higher frequency energy and pitch contour were preserved during resynthesis such that they matched that of the original /ɛ/ token. The resulting 12 step continuum therefore varied only in the frequencies of the first three formants. The onset /m/ from the original production of /mɛv/ was then re-spliced onto each continuum. The coda /b/ and /v/ were cross-spliced from productions of /mɛb/ and /mæv/ respectively. This was done to remove any possible acoustic traces of co-articulatory information from the preceding vowel cuing these consonants; though note it is unlikely that the cross-spliced stop closure/release and fricative noise contained cues to identify the original preceding vowel. Specifically, given that we predicted a following /v/ should bias listeners towards /ɛ/ categorization, as outlined above, the cross-spliced /v/ came from a post -/æ/ context, ensuring any possible acoustic information from the preceding vowel would predict the opposite adjustment in categorization. For the same reason /b/ was cross-spliced from a post-/ɛ/ context. These manipulations created 24 unique stimuli (12 continuum steps × 2 consonant frames).

2.4 Participants

Thirty-five self-identified native speakers of American English with normal hearing participated in Experiment 1. Participants were students at a North American Univerisity and received course credit for participation.

2.5 Procedure

Participants completed the task seated in front of a desktop computer, in a sound-attenuated booth in the lab. Stimuli were presented binaurally via a 3M™ Peltor™ listen-only headset. Participants were told that they would hear a speaker of English say nonce words, and that their task was simply to select which word they heard.

During a trial, participants were presented visually with two letters placed on either side of the computer screen: 'E' and 'A'. Prior to the trials beginning, participants were instructed that they should select 'E' if they heard the sound /ɛ/, and 'A' if they heard the sound /æ/. This was conveyed by giving examples of real words that rhymed with of the non-word continuum endpoints in the written computer instructions. Participants indicated their response by keypress, where an 'f' key-press indicated the letter on the left side of the screen and a 'j' keypress indicated a letter on the right side of the screen. The side of the screen on which each letter appeared was counterbalanced across participants. Participants completed 8 practice trials in which they heard each continuum end point in each CVC frame two times. During test trials participants heard each unique stimulus 8 times for a total of 192 trials. Stimuli were completely randomized. Testing took about 15 minutes.

2.6 Results & Discussion

Results were analyzed using Bayesian mixed-effects logistic regression, with the *brms* package (Bürkner, 2018) in R. We predicted the log odds of selecting an /ɛ/ response as a function of the step of the continuum, the consonantal frame (manipulating BP), and the interaction of these two fixed effects. Continuum step was treated as a continuous variable, and was scaled and centered. Consonantal frame was coded, with /mVb/ mapped to -0.5 and /mVv/ mapped to 0.5. Random effects in the model included by-participant intercepts with maximally specified random slopes including both fixed effects and their interaction. We employed weak normally distributed priors for both the intercept and for fixed effects, in both cases normal(0,1.5) (in log-odds space). In describing the results, we report the model estimates and 95% credible intervals for them. An effect is taken to have a reliable impact on responses when the

95% credible interval for an estimate excludes zero. We additionally computed a measure of the effects' distributions using the p_direction function in the package bayestestR (Makowski et al., 2019). This measure indexes the percentage of the posterior for an effect which shows a given sign, and ranges between 50% and 100%, if 99% of a given posterior is estimated to be positive, this would constitute relatively strong evidence for an effect with that directionality, and in this sense the probability of direction (pd) can be more intuitively compared to a frequentist p-value, to which it corresponds fairly closely. We would report the above case as pd= 99%. All model code and data for the experiments reported here is accessible through the OSF at https://osf.io/eba2v/.

In the model, the main effect of step was credible as expected ($\beta = 2.39$, CI = [1.80, 2.99]; pd = 100%). confirming that listeners /ɛ/ responses increased along the continuum. The main effect of consonantal frame was also credible ($\beta = 0.62$, CI = [0.04, 1.21]; pd = 98%). The estimate for the interaction between step and continuum did not provide credible evidence for an interaction: ($\beta = -0.12$, CI = [-0.34, 0.09]; pd= 87%). The effect of frame indicates that, consistent with biphone probability effects, participants showed an overall bias to categorize the target as /ɛ/ in the /mVv/ frame compared to the /mVb/ frame. This can be clearly seen in Figure 1.

FIG. 1 HERE

We can see qualitatively this effect involved a vertical shift in the categorization function not restricted to ambiguous regions of the continuum. Vertical shifts in categorization function have typically been attributed to decision biases (e.g. Massaro & Cowan, 1993, Norris et al., 2000). Contextual effects that directly modify input representations are predicted, instead, to only influence categorization at more ambiguous steps on a continuum (e.g. Massaro, 1989, Massaro & Cowan, 1993). Whether listeners use biphone probability information online and early in processing though, remains unclear. We tested this in Experiment 3. Irrespective of the nature of its influence, the results of Experiment 1 show that biphone probability can indeed modulate listeners' categorization of phonetic continua, in agreement with Pitt & McQueen (1998). Crucially, these results cannot be attributed to differences in neighborhood densities because we carefully controlled them during the stimulus selection.

3 Experiment 2: Neighborhood density effects controlling for biphone probability

Experiment 1 showed a clear effect of biphone probability in phonetic categorization of a non-word continuum, which was independent of neighborhood density. In Experiment 2 we tested if we could obtain evidence for an independent effect of neighborhood density. Two new continua were created: /bɛp/ ~/bæp/ and /bɛb/ ~ /bæb/. V_2C_3 biphone probability was matched completely for these two pairs (see Table 1), such that they both exhibited an equal /ɛ/ bias (-0.0019). Unlike Experiment 1 however, the neighborhood density bias for these continua was substantially different, both exhibited an /æ/ bias, though the bias for the /bɛp/ ~/bæp/ continuum (-29.35) was stronger than that for the /bɛb/ ~ /bæb/ continuum (-19.85). As outlined above, a denser neighborhood should bias listeners towards /æ/ responses, predicting that more /æ/ responses should be observed for the /bɛp/ ~/bæp/ continuum. Such a finding could not be explained by differences in biphone probability, which are identical (see Table 1). The empirical prediction is thus that a coda /b/ frame should show *increased* /ɛ/ responses (decreased /æ/ responses), as ND differences favor /æ/ more strongly in the frame with coda /p/. Here we note that the Breiss phonotactic probability calculator differs slightly from the KU phonotactic probability calculator, in showing a very small bias difference with coda /p/ slightly favoring /ɛ/.

3.1 Materials

Experiment 2 used the same vowel continuum created in Experiment 1, however, we presented them in different frames. The new frame consonants were cross-spliced from the same speakers' productions. The initial /b/ was cross-spliced from a production of /bɛb/. The coda /b/ was cross-spliced from a production of /bæb/, and the coda /p/ was cross-spliced from a production of /bɛp/. As with

Experiment 1, this method of cross splicing was chosen to remove any possible acoustic traces of the preceding vowel on cross-spliced coda consonants. Specifically, because we predicted that the /bæp/~/bɛp/ continuum should bias categorization towards /æ/ (as compared to /bæb/~/bɛb/), the coda /p/ was cross-spliced from an original /ɛp/ sequence. Likewise, the coda /b/ was cross-spliced from an original /æb/ sequence. Because the consonants used in Experiment 1 also involved labial constrictions, formant transitions at the onset and offset of the vowel continuum were judged to sound natural in these new frames.

3.2 Participants and procedure

Thirty-five self-identified native speakers of American English participated in Experiment 2. Participants were students at a North American University and received course credit for participation. The procedure for Experiment 2 was identical to that in Experiment 1.

3.4 Results & Discussion

The model specifications and model fitting procedure were identical to that in Experiment 1. Results are plotted in Figure 2. In contrast coding consonant frame, /bVp/ was mapped to -0.5 and /bVb/ was mapped to 0.5. As in Experiment 1, an expected main effect of step was credible ($\beta = 3.48$, CI = [2.59, 4.35]; pd = 100%). The main effect of consonant frame was also credible ($\beta = 3.84$, CI = [0.08, 0.59]; pd = 99.5%). In Figure 2 we can see the effect of consonant frame: consistent with predicted neighborhood density effects, listeners showed *increased* /ɛ/ responses with the /bVb/ frame, shifting categorization in accordance with neighborhood density. Unlike Experiment 1, there was evidence for an interaction between step and continuum ($\beta = 0.36$, CI = [0.00, 0.82]; pd = 97%), indicating that at higher steps on the continuum there is a more robust frame effect. These results replicate Newman et al.'s findings that neighborhood density affects phonetic categorization, and crucially preclude a biphone probability difference as a possible confound.

FIG. 2 HERE

4 Experiment 3: Time course of biphone probability and neighborhood density effects

Taking Experiment 1 and 2 together, we have evidence for an independent influence of both biphone probability and neighborhood density as indexed by listeners' categorization responses. However, categorization performance only provides a measure of the endpoint of the speech recognition process. To obtain precise timing information about when BP and ND effect recognition, we need evidence from online tasks. Previous research, outlined below, offers some relevant time course comparisons.

Using brain imaging, Pylkkänen, Stringfellow, & Marantz (2002) provide some evidence that biphone probability effects are consistently observed between 300 and 400ms post stimulus onset. In an MEG experiment, they administered a lexical decision task to listeners who were presented with CVC sequences that were either high probability and high density or low probability and low density. They investigated an MEG response component - M350 - which peaks between 300 and 400 ms post stimulus onset. Because the M350 was facilitated in response to the manipulated probability, and not inhibited as expected for a density manipulation, Pylkkänen et al argue that the M350 is sensitive to biphone probability. They did not find a clear correlate of the density effect in later MEG components. Thus, the MEG results present an estimate of the timeline for probability effects, and indirect support that this may be different from the effect of neighborhood density (see also Pylkkänen & Marantz, 2003),

More recently, Kingston and colleagues (Kingston, Levy, Rysling, & Staub, 2016) report on two experiments where they evaluated the time course of lexical effects on phonetic processing. In Kingston et al.'s experiments, listeners were asked to categorize a word to nonword phonetic continuum. They reasoned that if lexical effects are driven by feedback, they should be delayed as demonstrated in TRACE simulations (McClelland & Elman 1986). However, a rapid use of lexical information in categorization would constitute as evidence against feedback, and be more consistent with a feed-forward account.

Based on results from two eye-tracking experiments Kingston et al., claim that lexical effects influence phonetic processing between 300 and 400 ms after stimulus onset; and thus, are too early to be consistent with feedback.

A closer look at Kingston et al.'s experiments, however, offers an alternative explanation for their findings. First, in Kingston et al.'s Experiment 2a – the lexical effect is confounded with a biphone probability effect. In this experiment, listeners were presented with a continuum ranging between the vowels /ɛ/ and /ʌ/ in a CVC(C) frame; whether the end point was a word or non-word was determined by the final consonant. The continuum was placed in one of four frames: (1) /b _ ŋk/ forming the word "bunk" with /ʌ/, (2) /d _ ŋk/ forming the word "dunk" with /ʌ/, (3) /b _ ʃ/ and (4) /d _ ʃ/ (both resulting in nonwords). The initial consonant was varied to manipulate spectral context, and will not be discussed here; its inclusion does not alter the conclusions based on biphone probability differences discussed below. Because a coda /ŋk/ creates words with the vowel /ʌ/, but not /ɛ/, the Kingston et al. predicted that /ŋk/ should increase looks to an orthographic representation of /ʌ/ ("U"), as compared to a following /ʃ/. This is what the authors found, with the influence of the coda consonant(s) emerging within 300-400 ms of stimulus onset. A different interpretation of these finding emerges if we compare the biphone probabilities for the vowel and following consonant sequence. In the /ʃ/ context, the biphone probabilities are essentially matched with a very slight /ʌ/ bias: 0.0002 for /Cɛʃ/ and 0.0004 /Cʌʃ/. However, the biphone probability for the vowel and following consonant /ŋ/, reveals an asymmetry: a following /ŋ/ engenders a strong /ʌ/ bias: 0.0003 for /Cɛŋ/ and 0.0027 for /Cʌŋ/. The magnitude of this /ʌ/ bias is comparable to our own biphone probability manipulation in Experiment 1. Thus, an alternate explanation for Kingston et al.'s results is that the time course from Experiment 2a reflects a difference in biphone probability between the sequences, and therefore, like in Pylkkänen et al.'s MEG experiment, is observed between 300 and 400 ms post stimulus onset.

In the other eye tracking experiment reported by Kingston et al. (Experiment 1a), listeners categorized a continuum of fricative noise that ranged from /s/ to /f/. The continuum was followed by one of three frames: (1) /_ aɪl /,creating a word with /f/ "file", but not with /s/, (2) /_ aɪd /, creating a word with /s/ "side", but not with /f/, and (3) control frame /_ aɪm / for which both continuum endpoints were non-words. The online effect was significant only in the /_ aɪl / frame, with increased looks to a visual 'F' target on the screen, in comparison to the control frame. This effect cannot be explained by biphone probability differences; the summed biphone probability of "file" (0.0043) is slightly lower than that of "sile" (0.0058). However, there was no significant difference in looks between the /_ aɪm / frame and the control frame /_ aɪd /, where we would expect to see more looks to a visual 'S' target when the lexical context "side" reinforces /s/. This asymmetry in online processing between the two experimental frames makes it difficult to interpret the results from Kingston et al.'s Experiment 1a.

In Experiment 3 we used Kingston et al.'s experimental design with the stimuli used in Experiments 1 and 2, where the effects of biphone probability and neighborhood density were orthogonally manipulated. Specifically, we were interested in how these effects unfold online using a visual world eye-tracking task. Combining categorization with eye-tracking data allowed us to investigate the pre-decision stage integration of information as speech unfolds (unlike reaction times), as discussed in e.g. Norris et al. (2000). The eye movement response to the vowel spectra served as our baseline because it indexes a response to the signal. Given the independence of biphone probability and neighborhood density effects documented in Experiments 1 and 2 respectively, we expected to see an independent influence of each variable in the online task as well. If biphone probability effects are prelexical, we expected them to emerge soon after the spectral response (once listeners have heard the coda consonant). Of crucial interest was the relative timing of each effect. If neighborhood density effects originate from a feedback loop between the lexicon and prelexical information, because feedback takes time as shown in TRACE simulations (McClelland and Ellman 1986), the influence of ND should be delayed in comparison to a spectral response. Recall that Newman et al. (1997) reported reliable ND effects only at slow and intermediate reaction times, suggesting a later influence in processing.

4.1 Materials

The materials used in Experiment 3 were a subset of those used in Experiments 1 and 2. In order to present listeners with relatively ambiguous stimulus tokens (following e.g., Mitterer & Reinisch 2013, Reinisch & Sjerps, 2013), we presented listeners the most ambiguous region of each continuum. This was identified as the 4 step window centered around the 50% crossover points in the interpolated categorization functions derived from Experiment 1 and 2. In both experiments, categorization was most variable on steps 4 through 7. Participants heard all four continua (/mVb/, /mVv/, /bVb/, /bVp/) at these four steps. There were thus 16 unique stimuli used in Experiment 3 (4 continuum steps × 4 consonant frames).

4.2 Participants

Sixty-eight self-identified native speakers of American English with normal or corrected to normal vision participated in Experiment 3. We subsequently excluded three participants whose gaze data was not recorded consistently due to technical issues. We thus analyzed responses from sixty-five participants. Participants were students at a North American University and received course credit for participation.

4.3 Procedure

In Experiment 3, we used a visual world eye-tracking task, with a similar design to that used by Kingston et al. (2016). Participants were seated in front of an arm-mounted SR Eyelink 1000 (SR Research, Mississauga, Canada), which was set to track the left eye remotely, at a sampling rate of 500 Hz, and at a distance of approximately 550 mm. The visual display was presented to participants on a 1920 × 1080 ASUS HDMI monitor. Participants were tested in a sound-attenuated room in the lab. Participants' gaze was calibrated using a 5-point calibration procedure at the start of each experiment.

During an experimental trial, participants were presented with orthographic E and A on the target screen (Kingston et al. 2016) and were instructed to click on the letter corresponding to sound they heard. As in Experiments 1 and 2, examples of real English words that rhymed with the nonwords were given to convey the intended letter-to-sound mapping. Participants' eye movements were monitored while they performed the task. The orthographic targets were arranged vertically in the visual display, with each letter centered horizontally, and positioned 270 pixels above and below the midpoint of the display. Each letter was presented in 60pt black Arial font. The location of each letter was counterbalanced across participants. Each trial began with the appearance of a black fixation cross in the center of the visual display (60 px by 60 px). Following Kingston et al. (2016), stimulus onset was look-contingent, such that the audio stimulus played only after a look was registered on the fixation cross. Eye-movements were recorded from the first appearance of the fixation cross until a click response was registered by participants. After a click response was provided, the location of the mouse cursor was re-centered on the screen. Each trial was separated by a 1 second interval.

During the experiment, participants heard 8 repetitions of the 16 unique stimuli in a random order, for a total of 128 trials. Participants additionally completed 8 training trials prior to test trials in which they heard step 4 and step 7 for each frame, to give them practice with the experimental paradigm. The experiment took approximately 20 minutes to complete in total.

4.4. Analysis

We report several analyses of the Experiment 3 data. First, we analyzed listeners' click responses; we used Bayesian mixed-effects to model the log odds of selecting an /ɛ/ response as a function of frames (/mVb/, /mVv/, /bVb/, /bVp/), as in Experiments 1 and 2. The model was fit with the same fixed effects and random effect structure as previous models.

We additionally carried out two complementary eyetracking analyses. First, we report results from the traditional moving-window analysis for which the time series data were binned in 100 ms intervals. We report this for ease of comparison to findings from Kingston et al. (2016). One shortcoming of moving window analyses however, is their inability to capture the inter-relatedness of adjacent time bins. A binned analysis is further non-ideal for model changes over time and trajectories which are non-linear (as is often the case in eye movement data). To address these limitations, we also report results from analysis using a Generalized Additive Mixed Model (GAMM), which offers a powerful tool for analyzing time-series data from visual world experiments (Nixon, van Rij, Mok, Baayen & Chen, 2016; Zahner, Kutscheid & Braun, 2019; Steffman 2021). GAMMs have recently been advocated for use in modeling eye movement data, as they (1) easily fit non-linear trajectory shapes, (2) provide for an intuitive assessment of when eyetracking trajectories diverge (see Zahner et al. 2019 for similar discussion advocating for GAMMs).

The dependent measure for both analyses was the same: a "preference" measure computed as listeners' log-transformed looks to /ɛ/ subtracted from their log-transformed looks to /æ/ (see e.g., Reinisch & Sjerps, 2013). Each measure was transformed using the empirical logit (Elog) transformation, as described in Barr (2008). For both analyses, the analysis window was 0 to 1200 ms after the onset of the target vowel; listeners typically made a categorization click response within this time period, after which there was a substantial drop in recorded eye-movements.

In the moving window analysis, with the data binned into 100 ms intervals, a separate linear mixed effects regression was run, again using *brms* fit with weak normal priors. In each binned regression, the dependent measure was predicted as a function of continuum step (scaled), and frame, which in this case we contrast coded. We subsequently extracted two estimates of pairwise frame differences of interest, using emmeans (Lenth, 2020). These were: /mVb/ versus /mVv/ (indexing the BP effect), and /bVb/ versus /bVp/ (indexing the ND effect). The estimate and distribution for each marginal comparison was then computed,, in addition to the effect of continuum step.

The GAMM model was implemented with the *mgcv* and *itsadug* packages in R (van Rij, Wieling, Baayen, & van Rijn, 2020; Wood 2011). We implemented an AR1 error model, following procedures described in Sóskuthy (2017), which was found to greatly reduce residual autocorrelation inherent in timeseries data (see open access model code for implementation). The numerical model output is fairly uninformative for understanding the timing questions asked here (Wood 2011; Zahner et al., 2019), but the summary is included in Appendix 2 for completeness. The dependent variable in the model was the same log-transformed preference measure, though in this analysis it was binned in 20 ms intervals and thus provides a more fine-grained comparison of timing effects (as in Zahner et al., 2019; Steffman, 2021). The model included parametric terms for consonant frame. GAMMs are also the preferred way to analyze eye tracking data because they involve non-linear smooths that better approximate the dynamics of fixation data. Smooth terms in the model were fit to model a non-linear interaction of (scaled) continuum step by frame over time and including the main effects of smooths as well (using the *te( )* function, cf. Nixon et al. 2016). By-participant random smooths over time (factor smooths), as well as factor smooths by consonant frame, analogous to by-participant intercepts and slopes for frame in mixed models, were also included (see e.g., Sóskuthy 2021). For both of the random effect (factor smooth) terms, the *m* parameter was set to 1 (Baayen, van Rij, de Cat, & Wood 2018). The numerical model output is in the Appendix, and the estimates are presented in Figure 6.

4.5 Results & Discussion

4.5.1 Click responses

FIG. 3 HERE

Overall, in the Experiment 3, the continuum steps we used (4-7) were perceived as more /æ/-like; this is evident from the credible negative intercept estimate for the reference level which was set to be /mVb/ (β = -0.41, CI = [-0.68,-0.15]; pd = 100%) as shown in Figure 3. This /æ/ bias was stronger than in Experiments 1 and 2. It is possible that despite sampling from around the 50% crossover point, listeners recalibrated categorization because of the absence of steps from the endpoints of the continua. Nevertheless continuum step still showed a credible effect as expected. Listeners increased /ɛ/-responses (β = 1.26, CI = [1.02, 1.53]; pd = 100%) progressively from Step 4 towards Step 7 where formants were more /ɛ/-like. Thus, although listeners exhibited an overall /æ/ bias, they still used formant cues in the expected way.

The first comparison of interest was between the frames manipulating biphone probability: /mVb/ versus /mVv/; with /mVb/, as the reference level in the model, the estimate for the /mVv/ frame was credibly positive (β = 0.47, CI = [0.04, 0.89]; pd = 98%), replicating the observed difference between these two frames in Experiment 1. There was also evidence for an interaction between continuum step and the /mVv/ frame: (β = 0.29, CI = [0.06, 0.53]; pd = 99.4%), showing that the effect of biphone probability was larger at higher continuum steps. No other interactions between the two other frames and step were credible in the model.

The model estimates showed further that both /bVb/ and /bVp/ frames evidenced credibly decreased /ɛ/ responses relative to the /mVb/ (/bVb/: β = -1.24, CI = [-1.63, -0.84]; pd = 100%; /bVp/: β = -1.28, CI = [-1.71, -0.87]; pd = 100%). In fact, this difference in /æ/ responses between the /m/- vs /b/-initial frames was even larger than the biphone probability effect across the /m/-initial frames. Pairwise comparisons between /mVv/ and both /b/-initial frames were examined using emmeans (Lenth, 2020), and as expected based on the Figure 3, were each credibly different from one another.

Before we turn to the comparison between /b/-initial frames, let us consider the difference we see here based on initial consonant. This effect emerged in Experiment 3 because we used a within-subject design in contrast to the between-subjects design in Experiments 1 and 2 where the effects of the /m/-

initial and /b/-initial frames were investigated separately. We can be sure that this effect was not driven by the differences in biphone probabilities of the /m/-initial and /b/-initial frames. From Table 1 (using the Vitevitch & Luce metrics) we see that the biphone sequence /mV/ has a stronger /æ/ bias (-0.0042) compared to /bV/ (-0.0027); this difference in biphone probability would predict the opposite of the effect observed here. Even considering the summed biphone probability of the whole CVC sequence, we see the following gradation in the strength of /æ/ biases, from largest to smallest: /m_b/ (-0.0061) > /b_p/ and /b_b/ (-0.0046) > /m_v/ (-0.0035). This too cannot explain the difference we see between /m/- and /b/-initial frames, because based on this rank ordering the most /æ/ responses are expected for /m_b/, which is clearly not the case.

The direction of difference in /ɛ/ responses between the the /m/- and /b/-initial frames is more consistent with a difference in neighborhood density with the latter having a stronger /æ/ bias (Table 1). However, there is also reason to be skeptical that neighborhood density differences are driving the difference between /m/- and /b/-initial frames. The difference in neighborhood density between the two /b/-initial frames was at least as large, if not larger in magnitude than the neighborhood density difference between the /m/- and the /b/-initial frames. Yet, the effect between /m/- and /b/-initial frames was credible, but (not) the neighborhood density effect indexed by the difference between the two /b/-initial frames (reported below).

Instead, we speculate that by introducing different initial consonants in our frames, we may have introduced a new variable that influenced listeners' perception of the target vowel. A change in initial-consonant from /b/ to /m/ is a switch between an oral and a nasal onset. Although our vowel didn't vary in terms of nasality across frames (being originally produced in /m/ initial frames), listeners' perception of F1 and/or F2 is likely to have been modulated because they were compensating for the typical coarticulatory effects of nasals on vowel formants. Nasalization of vowels adjacent to nasal consonants is well-attested in American English (e.g. Chen, Slifka & Stevens 2007; Cohn 1990). Nasalization typically lowers perceived F1 for low vowels (Diehl, Kleunder & Walsh, 1990), directly impacting listeners' perception of vowel height adjacent to nasal consonants (Beddor, 1993; Ohala, Beddor, Krakow & Goldstein 1986; Wright 1980). Ohala et al. (1986) present a test case that offers a close comparison to the present stimuli. They found that when a vowel on an /ɛ/ ~ /æ/ continuum was adjacent to a nasal consonant, but had only very weak nasalization (comparable to the present stimuli where vowels were originally produced in /m/-initial carryover contexts) listeners "overcompensated" for the expected effect of vowel nasalization. An adjacent nasal consonant accordingly lead to decreased /æ/ responses, i.e. perception of a higher vowel, /ɛ/. Thus, it is quite likely that the difference between /m/- and /b/-initial frames is due to the listener's compensation for the nasal context, and not attributable to either biphone probability or neighborhood density differences. We addressed this issue directly in Experiment 4.

The second comparison of interest was between the frames manipulating neighborhood density: /bVb/ versus /bVp/, also extracted using emmeans. Unlike in Experiment 2, there was no credible difference between these two frames used to manipulate neighborhood density ($\beta = 0.05$, CI = [-0.34, 0.26]). As we can see from Figure 3, these frames did not induce any reliable shift in categorization. Thus, we did not replicate the neighborhood density effect observed in Experiment 2. We return to this point in discussing the eye movement data below. Given that the results of Experiment 3 diverge from Experiment 2, we note here that we replicated the categorization task from Experiment 3 in Experiment 5, which is not contained in the paper, but for which the code, and summary of the results, is available on the open access repository.

4.5.2 Eye Movement data

In Figure 4, we plot listeners' proportion of looks to /ɛ/ over time (not the log transformed preference measure used in modeling) for ease of visual inspection. In this figure the time course of looks to /ɛ/, split by consonant frame (panel A), continuum step (panel B), and frame faceted by step (panel C) are presented. First, confirming what we saw in the categorization responses, it is clear the eye movement data show a strong bias towards /æ/, that is, listeners' fixations to /ɛ/ are overall fairly low. Qualitatively, we can note that the frame effects shown in Figure 4 panel A mirror the categorization responses described in section 4.5.1: there is a clear separation between the BP-manipulating frames, with /mVv/ favoring looks to /ɛ/, in contrast to theND-manipulating frames, which are overlapping. As with the categorization results, we additionally see a robust effect of initial consonant, with /m/-initial frames favoring looks to /ɛ/.

FIG. 4 HERE

In panel B, we can see that continuum step exerted an expected influence in online processing: higher values (more /ɛ/-like steps) favor looks to /ɛ/. Finally, in panel C we can see that there are differences in the timing and magnitude of the frame effects based on continuum step. Each of these results is discussed in detail below.

4.5.2.1. Moving window analysis

FIG. 5 HERE.

Full model outputs and code for running the models can be obtained from the open access materials hosted on the OSF (https://osf.io/eba2v/). In reporting the results of the moving window analysis, we focus on summary statistics for each of the three estimates and their 95% credible intervals over (binned) time (Figure 5). The three estimates plotted are for the influence of continuum step, and the pairwise comparison between /mVb/ to /mVv/ frames – the biphone effect, and that of the /bVb/ to /bVp/ frames – the neighborhood density effect. As with the categorization results, when CI excludes zero, we can be confident that there is a robust influence of a given manipulation on eye movements. The online OSF repository also includes a visualization of this same time course with the proportion of direction measure.

In considering the timing of the effects, the effect of continuum step provides a baseline capturing listeners' rapid use of spectral cues in processing (Reinisch & Sjerps 2013). As shown in Figure 5, we can see that estimates for continuum step reliably exclude zero for the 400-500 millisecond bin in the time series. That is, listeners' responses to the vowel continuum were reliable within 400-500ms after the target vowel onset. This is slightly slower than previous reports for the use of intrinsic spectral cues; for example, Kingston et al. found a reliable effect of vowel acoustic in the 300-400 time bin window in their analysis (cf. Reinisch & Sjerps, 2013). We attribute this delay to listeners' possible reliance on vowel duration as a cue to the /ɛ/-/æ/ contrast. In our experiment the duration of the vowel was also longer (260 ms) compared to that in previous studies (approximately 150 ms in the case of Kingston et al.). The delay could also be driven in part by the biased nature of the continuum. Regardless of the reasons for the discrepancy, the timing for use of vowel-intrinsic spectral cues provides a baseline for evaluating the effects for frames of interest.

First, turning to the effect of ND-manipulating frames /bVb/ versus /bVp/, we can see that at no point does the distribution for the estimate of this effect exclude zero. This replicates the lack of an ND effect in the categorization responses and is also consistent with the visual inspection of the eyetracking data described above. There is some weak evidence for a delayed difference between these two frames in the 1000-1100 ms bin, because 90% of the CI in this analysis window excludes 0.

Turning to the effect of BP-manipulating frames /mVb/ versus /mVv/, we can see that clear evidence for an effect of BP emerges within 500-600 ms after the target vowel onset, and remains robust throughout the remainder of the analysis window. Consider again that because the vowel is 260 ms in duration, information about the coda consonant is available only at that point. Given that it takes approximately 200 ms to initiate a saccade (Dahan, Magnuson, Tanenhaus & Hogan, 2001; Matin, Shao & Boff 1993), this effect's timing suggests listeners fairly rapidly integrated coda consonant information information with their perception of the vowel (the earliest possible timing for this effect would be in the 400-500 ms bin). Thus, we have evidence for a robust effect of BP which closely follows the effect of vowel acoustics in time, indicating a rapid effect. Note that the absolute value of the timing of the effect in this experiment is about 100 ms longer than that reported in Kingston et al., (2016), which is consistent with the difference in vowel duration between our stimuli and theirs (260ms here compared to 170ms in Kingston et al., Experiment 2a).

Though not a focus of interest here, we can note that the effect of initial consonant was also robust and early, as assessed in the moving window analysis. The pairwise difference between /m/- and /b/-initial frames were credible even in the 200-300 ms window, that is, even before the vowel effect, as might be expected for effects relating to the initial consonant. Such early effects are unlikely to be related to neighborhood density. We used a GAMM analysis to better characterize the relationship between continuum step and the frame effect, and to offer a different assessment of the timing of the BP effect.

4.5.1.2. GAMM analysis

FIG. 6 HERE

Because the GAMM analysis provides more fine-grained information about the time course (bins are 20ms not 100ms) and takes into account the relationship between adjacent time bins, we used it to evaluate the interaction between continuum step and the BP and ND effects. We expected only early, prelexical effects to interact with the bottom-up information in the signal as exemplified by the continuum steps. To assess the extent to which continuum step and consonant frame interacted in our GAMM analysis, we compared our model fit with the interacting *te()* term to one in which step and frame each had separate smooths which did not interact, using the *compare_ML()* function in *itsadug*. The model allowing for an interaction between continuum step and consonant frame provided a better fit to the data ($\chi^2(9) = 24.16$, $p < 0.001$; see https://osf.io/eba2v/ for the full code for model comparison).

In Figure 6 we plot the difference smooths between consonant frames of interest, that is, comparing /mVb/ to /mVv/ - the BP effect, and /bVb/ to /bVp/ - the ND effect at each continuum step. These model estimates represent the *difference* between two smooths, with confidence intervals. The time when this estimated difference reliably becomes different than zero, i.e., when the confidence intervals for the estimate exclude zero, is when an effect is taken to be reliable (see e.g., Zahner et al, 2019; Steffman 2021).

As shown in panel A of Figure 6, we see a robust divergence from zero at all continuum steps for the BP effect arising from the comparison between the /mVb/ and /mVv/ frames. First, there was a clear relationship between continuum step (vowel acoustics) and the timing of the effect. Specifically, the biphone probability information was more rapidly integrated when vowel information was more /ɛ/-like (Step 6 & 7), than when it was /æ/-like. The sensitivity of the BP effect to fine-grained differences in vowel acoustics is consistent with the claim that it is an early influence in processing. In the context of an /æ/ biased experiment, acoustic evidence for /ɛ/ would support listeners' integration of /ɛ/ with the coda consonant favoring a high biphone probability sequence: in other words, when both the vowel acoustics and consonant frame favor /ɛ/, divergence occurs more quickly. This effect was as early as 473 ms (range 473 -703) from target vowel onset. In contrast, the neighborhood density effect represented by the difference smooth comparing /b_b/ and /b_p/ frames in Figure 6, panel B did not diverge from 0 at any point in the analysis window. That is, we did not observe an ND effect online, lining up with listeners' click responses, and the moving window analysis.

In summary, the GAMM analysis allows us to confirm (1) a robust and rapid influence of biphone probability in online processing, and (2) a lack of a robust influence of neighborhood density. Both of these conclusions line up with our moving window analysis. We further saw that vowel acoustics influence the dynamics of processing for BP information, that is, acoustic support for /ɛ/ (in an overall -/æ/-biased experiment) leads to an earlier influence of biphone probability.

In Experiment 4, we probed the unexpected difference between the /m/- and /b/-initial frames further, testing the claim that this effect is not attributable to BP or ND differences.

5 Experiment 4

Recall that the stronger /æ/ bias for /b/-initial frames observed in Experiment 3 was consistent with the small neighborhood density difference favoring /b/-initial compared to /m/-initial frames. However, its early timing as well as the difference in magnitude of the effect compared to the neighborhood density effect observed in Experiment 2 led us to hypothesize that this effect was not driven by the neighborhood density differences. Instead, we conjectured that the frame effect was driven by perceptual adjustments related to nasal consonants and their effects on judgements of vowel height. Experiment 4 was designed to confirm that the difference between /m/- and /b/-initial frames seen in Experiment 3 was unrelated to neighborhood density and biphone probability. In Experiment 4 we presented listeners with another /m/-initial and /b/-initial frame where *both* biphone probability and neighborhood density predicted the opposite of the observed difference between /m/- and /b/-initial frames seen in Experiment 3. If we replicated the nasal vs oral frame effect from Experiment 3 here, we could be sure that it was not driven by either biphone probability or neighborhood density differences.

5.1 Materials

The frames used in Experiment 4 were /m_v/ (used in Experiment 1) and /b_v/. To create the new /b_v/ frames, the initial /b/ from the continua used in Experiment 2 was cross spliced, replacing the /m/ in the /m_v/ frames. As shown in Table 1, both biphone probability and neighborhood density predict that an /b_v/ should show increased /ɛ/ responses relative to the /m_v/ frame. This is the opposite of the effect seen in Experiment 3 (where the /b/-initial frames showed *decreased* /ɛ/ responses), and accordingly, we can test if the effect observed there is independent of both biphone probability and neighborhood density.

5.2 Participants and procedure

Thirty-four self-identified monolingual English-speaking participants were recruited to participate in Experiment 4. Unlike previous experiments, these participants were recruited online, via the platform Prolific, and completed the experiment over the internet. Participants were instructed to complete the experiment seated in a quiet room with a pair of headphones. Participants were paid 4$ for this experiment which took 15-20 minutes to complete. The experimental procedure was otherwise identical to that in Experiments 1 and 2.

FIG. 7 HERE

5.3 Results and discussion

Listeners' categorization responses were assessed by the same method and model structure as used in previous experiments. In contrast coding the frames, /m_v/ was mapped to -0.5 and /b_v/ was mapped to 0.5. Continuum step had a credible effect on responses, as seen in all previous experiments ($\beta = 3.65$, CI = [2.99,4.30]; pd = 100%). Additionally, consonant frame had a credible effect ($\beta = -0.68$, CI = [-1.31,-0.08]; pd = 99%). Replicating the effect observed in Experiment 3, listeners showed decreased /ɛ/ responses for the /b_v/ frame, as shown in Figure 6. The interaction between frame and step was not credible ($\beta = -0.17$, CI = [-0.67,0.29]; pd = 76%). The direction of the effect of consonant frame in this experiment, despite opposing neighborhood density and biphone probability effects confirms that the robust difference between /m/-initial and /b/-initial frames in Experiment 3 was not driven by differences in neighborhood density (or biphone probability).

7 General discussion

In four experiments, we tested how differences in biphone probability and neighborhood density influence listeners' categorization of a vowel continuum embedded in nonwords. In Experiment 1, we found that listeners shifted categorization to form a high probability sequence even when stimuli were controlled for neighborhood density. Likewise, in Experiment 2, we found that listeners shifted categorization to favor a denser neighborhood even when stimuli were controlled for biphone probability. Finally, in Experiment 3, we found evidence for a robust and early influence of biphone probability. In contrast, density effects affected neither categorization nor looking behavior. In one additional experiment, we probed an unexpected influence uncovered in Experiment 3. This effect resulted from mixing the stimuli from Experiment 1 and 2, and was not driven by biphone probability or neighborhood density; instead, it was due to context effects related to a preceding nasal consonant.

Our results provide both direct and indirect evidence for a dissociation between biphone probability and neighborhood density effects. In Experiments 1 and 2 we showed that biphone probability and neighborhood density effects exert an independent influence on offline categorization. That is, despite the correlation between biphone probability and neighborhood density in English, biphone probability effects in phonetic processing cannot be explained by differences in neighborhood activation alone as we show in Experiment 1. Similarly, neighborhood density effects in phonetic processing can also not be explained by differences in biphone probabilities alone, as we show in Experiment 2.

Categorization data from Experiment 3 also provided evidence for a dissociation, albeit indirect. In Experiment 3, the mixing of stimuli from Experiments 1 and 2 increased the variability of frames (which had a clear effect on responses as confirmed in Experiment 4). Despite the inclusion of more variable frames in Experiment 3, the biphone probability effect on categorization replicated the effect observed in Experiment 1 with fewer frames. However, neighborhood density influences, which were fairly small in magnitude in Experiment 2, disappeared when an irrelevant dimension of variation (in the initial consonant) was introduced into Experiment 3. That is, biphone probability effects were robust across online and offline tasks, and not affected by the increased variability in Experiment 3. In comparison, the increased variability in frames and task complexity in Experiment 3 led listeners to disregard neighborhood density differences in the stimuli. Together, these categorization results are consistent only with accounts where both biphone probability and neighborhood density independently influence processing, albeit in qualitatively distinct ways.

Independent contributions of BP and ND effects seen here provide clear constraints on existing models of spoken word recognition. This is problematic for models like TRACE that do not independently represent biphone probability information (cf. Pitt & McQueen, 1996). This is also incompatible with Norris et al.'s (2000) proposal that neighborhood density effects are rooted in biphone probability differences, and therefore suggests that for models like Merge (Norris, 1999, Norris et al., 2000) and Shortlist (Norris, 1994), information from the lexicon must be used to model effects of ND on phonetic processing.

The categorization results from Experiment 3 also suggest that biphone probability and neighborhood density affect processing at different times. A general consensus in the literature is that early influences in processing are not impacted by task factors (e.g. Miller & Dexter 1988), including the presence of orthogonal variation in stimuli of the kind introduced in Experiment 3 (Green, Tomiak & Kuhl, 1997), as well as cognitive load (Bosker et al., 2017). Thus, based on robustness across tasks and stimulus variability, it is likely that biphone probability, but not neighborhood density, affects processing early.

The eye-tracking data from Experiment 3 confirmed that biphone probability effects are indeed early; biphone probability information is incorporated as early as 400-500 ms after the onset of the vowel, at the shortest just 210 ms after the onset of the BP-manipulating coda consonant. Further, when the vowel formants were more /ɛ/-like, biphone probability information was integrated the earliest in processing. The time course of the biphone probability effect in our experiments is similar to the timing of the effect in Kingston et al.'s (2016) findings. In their experiments, as well as in Experiment 3, biphone probability effects emerged less than 50ms into the coda consonant. Given that our biphone probability results cannot be attributed to lexical factors because our continuum endpoints were non-words, and neighborhood density was controlled, we take these aligning time course results to strengthen our argument that biphone probability differences are responsible for Kingston et al.'s findings in Experiment 2a.

The independence of the biphone probability effect and its early timing precludes biphone probability effects from being an epiphenomenon of lexical feedback (cf. Newman et al. 1997). Instead, the rapid, independent biphone probability effects observed here are consistent with proposals that biphone probability is encoded prelexically, and varies as a function of the robustness of the speech signal (Pitt & McQueen, 1998; Pylkkänen, Stringfellow, & Marantz, 2002; Norris et al. 2000).

While we were able to disassociate BP and ND effects, we were unable to delineate the time course of the ND effect. By virtue of its fragility, the increased variability introduced in the eye-tracking experiment (Experiment 3) due to the inclusion of multiple frames erased the small effect of ND observed in the offline task in Experiment 2. In the absence of precise timing information about ND effects, we cannot conclude whether ND effects are a consequence of feedforward activation to decision nodes, or due to feedback effects from the lexicon. However, based on the lack of robustness of ND effects, we can rule out the possibility that ND affects processing as early as BP. Future eye tracking experiments will be required to confirm if they are as delayed as might be expected if they are a result of feedback (Newman et al., 1997), or only moderately so as expected if they feedforward to decision nodes (Norris et al., 2000).

In conclusion, we present new evidence for the dissociation of biphone probability and neighborhood density effects using a combination of categorization and online processing measured with eye tracking. Our results offer support for the claim that biphone probability influences in perception are independent from that of neighborhood density, prelexical in nature, and operate early in processing. Based on these results we argue in favor of models that encode both biphone probability and neighborhood density, albeit with asynchronous timing effects on early processing. Further research will be needed to establish a precise time course for neighborhood density effects, and from extending the present findings to determine how they combine with other known influences, such as word-hood and word frequency.

References:

Archer, S. L., & Curtin, S. (2016). Nine-month-olds use frequency of onset clusters to segment novel words. *Journal of Experimental Child Psychology*, *148*, 131-141.

Barr, D. J. (2008). Analyzing 'visual world' eye tracking data using multilevel logistic regression. *Journal of memory and language*, *59*(4), 457-474.

Baayen, R H., R Piepenbrock, and L Gulikers. CELEX2 LDC96L14. Web Download. Philadelphia: Linguistic Data Consortium, 1995.

Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. (2018). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. In *Mixed-effects regression models in linguistics* (pp. 49-69). Springer, Cham.

Beddor, P. S. (1993). The perception of nasal vowels. In *Nasals, nasalization, and the velum* (pp. 171-196). Academic Press.

Bürkner P. (2018). "Advanced Bayesian Multilevel Modeling with the R Package brms." *The R Journal*, 10(1), 395–411.

Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2017). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language*, *94*, 166–176.

Breiss, C. (2021) Phonotactic Probability Calculator. Retrieved from https://osf.io/eba2v/.

Chen, N. F., Slifka, J. L., & Stevens, K. N. (2007). Vowel nasalization in American English: acoustic variability due to phonetic context. *Speech Communication*, 905-918.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507-534.

Diehl, R. L., Kluender, K. R., & Walsh, M. A. (1990). Some auditory bases of speech perception and production. *Advances in speech, hearing and language processing*, *1*, 243-268.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(4), 526–540.

Friederici, A. D., & Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & psychophysics*, *54*(3), 287-295.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of memory and language*, *42*(4), 481-496.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and language*, *45*(3), 468-492.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 84.

Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics*, *59*(5), 675–692.

Hayes, B. (2012). The role of computational modeling in the study of sound structure. Paper presented at Laboratory Phonology.

Hollich, G., Jusczyk, P. W., & Luce, P. A. (2002, November). Lexical neighborhood effects in 17-month-old word learning. In *Proceedings of the 26th annual Boston University conference on language development* (Vol. 1, pp. 314-23). Boston, MA: Cascadilla Press.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*(5), 630.

Kingston, J., Levy, J., Rysling, A., & Staub, A. (2016). Eye movement evidence for an immediate Ganong effect. *Journal of Experimental Psychology. Human Perception and Performance*, *42*(12), 1969–

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, *12*(2), 119–131.

Lenth, R. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.3. https://CRAN.R-project.org/package=emmeans

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*(1), 1.

Maeda, S. (1993). Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In *Nasals, nasalization, and the velum* (pp. 147-167). Academic Press.

Makowski D., Ben-Shachar M., Lüdecke D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software*, *4*(40), 1541.

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2020). Eye-tracking the time course of distal and global speech rate effects. *Journal of Experimental Psychology: Human Perception and Performance*. *46*(10), 1148-1163.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive psychology*, *21*(3), 398-421.

Massaro, D. W. & Cowan, N. (1993) Information processing models: Microscopes of the mind. Annual Review of Psychology 44:383–425

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & psychophysics*, *53*(4), 372-380.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78*(2), 91-121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive psychology*, *38*(4), 465-494.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1-86.

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 369.

Mitterer, H., & Reinisch, E. (2013). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*, *69*(4), 527-545.

Munson, B., Swenson, C. L., & Manthei, S. C. (2005). Lexical and phonological organization in children. *Journal of Speech, Language, and Hearing Research*.

Munson, B., Edwards, J., & Beckman, M. E. (2005). Phonological knowledge in typical and atypical speech–sound development. *Topics in language disorders*, *25*(3), 190-206.

Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(23), 873–889.

Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language*, *90*, 103-125.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, *52*, 189–

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–325.

Ohala, J. J., Beddor, P. S., Krakow, R. A., & Goldstein, L. M. (1986). Perceptual constraints and phonological change: a study of nasal vowel height. *Phonology*, *3*, 197-217.

Pierrehumbert, J. B., Needle, J., & Hay, J. B. (2018). Phonological and morphological effects in the acceptability of pseudowords (A. Sims & A. Ussishkin, Eds.). Cambridge University Press.

Pitt, M. A., & McQueen, J. M. (1998). Is Compensation for Coarticulation Mediated by the Lexicon? *Journal of Memory and Language*, *39*(3), 347–370.

Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and language*, *81*(1-3), 666-678.

Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, *41*(2), 101-116.

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 29.

Sebastián-Gallés, N., & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: Role of early exposure. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 974.

Sóskuthy, M. (2017). Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. *arXiv preprint arXiv:1703.05339*.

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 101017.

Steffman, J. (2021). Prosodic prominence effects in the processing of spectral cues. *Language, Cognition and Neuroscience*, *36*(5), 586-611.

Stevens, K.(1998). Acoustic phonetics. Cambridge, MA: MITPress.

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological science*, *13*(5), 480-484.

Thorn, A. S., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 729.

van Rij J., Wieling M., Baayen R., van Rijn H. (2020). itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs. R package version 2.4.

Vitevitch, M. S. (2002a). Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of the ear. *Language and speech*, *45*(4), 407-434.

Vitevitch, M. S. (2002b). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 514.

Vitevitch, M. S., & Luce, P. A. (1998). When Words Compete: Levels of Processing in Perception of Spoken Words. *Psychological Science*, *9*(4), 325–329. Retrieved from JSTOR.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language*, *40*(3), 374–408.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 481-487.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, Neighborhood Activation, and Lexical Access for Spoken Words. *Brain and Language*, *68*(1), 306–311.

Weide, R. L. (1998). The Carnegie Mellon pronouncing dictionary. *release 0.6, www.cs.cmu.edu*.

Winn, M. (2016). Praat script: Make formant continuum [Computer software]. Retrieved 15 January, 2018, from http://www.mattwinn.com/praat.html.

Wood SN (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, **73**(1), 3-36.

Wright, J. (1980). The behavior of nasalized vowels in the perceptual vowel space. *Report of the Phonology Laboratory Berkeley, California*, (5), 127-163.

Zahner, K., Kutscheid, S., & Braun, B. (2019). Alignment of f0 peak in different pitch accent types affects perception of metrical stress. *Journal of Phonetics*, *74*, 75-95.

Appendix 1: GAMM model output

TABLE 2 HERE

Table 1: Lexical statistics and biases for the continuum endpoints used in the experiments. See section 2 for details on calculation of biphone probability (BP) and neighborhood density (ND). Endpoint biases are calculated with /ɛ/ as reference; thus, positive numbers favor greater /ɛ/ responses. The absolute difference between endpoint responses is given below each endpoint pair in bold. Two different BP calculations are reported in two separate columns (see text).

Experiment 1	BP (Vitevich & Luce)		BP (Breiss)	ND (Vitevich & Luce)	
	C_1 V_2	V_2C_3	CVC	C_1 V_2	CVC
/mæb/	0.0101	0.0026	0.0121	31.95	29.54
/mɛb/	0.0059	0.0007	0.0069	31.06	17.96
bias (positive = /ɛ/)	*-0.0042*	*-0.0019*	*-0.0052*	*-0.89*	*-11.58*
/mæv/	0.0101	0.0019	0.0118	31.95	30.25
/mɛv/	0.0059	0.0026	0.0091	31.06	17.37
bias (positive = /ɛ/)	*-0.0042*	*0.007*	*-0.0027*	*-0.89*	*-12.88*
***bias* difference**	**matched**	**0.0026**	**0.0025**	**matched**	**matched**

/m_v/ favors /ɛ/ based on BP

Experiment 2	BP (Vitevich & Luce)		BP (Breiss)	ND (Vitevich & Luce)	
	C_1 V_2	V_2C_3	CVC	C_1 V_2	CVC
/bæb/	0.0059	0.0026	0.0089	46.84	41.11
/bɛb/	0.0032	0.0007	0.0098	33.47	21.26
bias (positive = /ɛ/)	*-0.0027*	*-0.0019*	*0.009*	*-13.37*	*-19.85*
/bæp/	0.0059	0.0048	0.0106	46.84	44.42
/bɛp/	0.0032	0.0029	0.0121	33.47	14.46
bias (positive = /ɛ/)	*-0.0027*	*-0.0019*	*0.0014*	*-13.37*	*-29.96*
***bias* difference**	**matched**	**matched**	***0.0005***	**matched**	**10.11**

/b_b/ favors /ɛ/ based on ND

Experiment 4	BP (Vitevich & Luce)		BP (Breiss)	ND (Vitevich & Luce)	
	C_1 V_2	V_2C_3	CVC	C_1 V_2	CVC
/bæv/	0.0059	0.0019	0.0085	46.84	24.74
/bɛv/	0.0032	0.0026	0.120	33.47	15.19
bias (positive = /ɛ/)	*-0.0027*	*0.007*	*0.0035*	*-13.37*	*-9.55*
***bias* diff. v.s. /mɛv/ ~ /mæv/**	***0.0015***	**matched**	***0.0062***	**12.48**	**-3.33**

/b_v/ favors /ɛ/ based on BP & ND

Table 2: Output for the GAMM, showing parametric and smooth terms. Note the reference level in the parametric terms in /m_b/.

Parametric terms	Est.	SE	t	p
Intercept	-0.72	0.08	-9.20	< 0.001
frame /b_b/	-0.18	0.05	-3.61	< 0.01
frame /b_p/	-0.18	0.05	-3.80	< 0.01
frame /m_v/	0.25	0.06	4.19	< 0.001
Smooth terms	edf	Ref. df	F	p
te(time, continuum):frame /b_b/	11.11	14.12	13.50	< 0.001
te(time, continuum):frame /b_p/	11.54	13.98	12.79	< 0.001
te(time, continuum):frame /m_b/	18.29	24.78	16.62	< 0.001
te(time, continuum):frame /m_v/	14.54	16.74	28.70	< 0.001
s(time, participant)	490.82	584.00	9.02	< 0.001
s(time, participant): frame.ordered: /b_b/	163.61	584.00	0.48	< 0.001
s(time,participant): frame.ordered: /b_p/	164.36	584.00	0.47	< 0.001
s(time,participant): frame.ordered: /m_v/	223.41	584.00	0.82	< 0.001

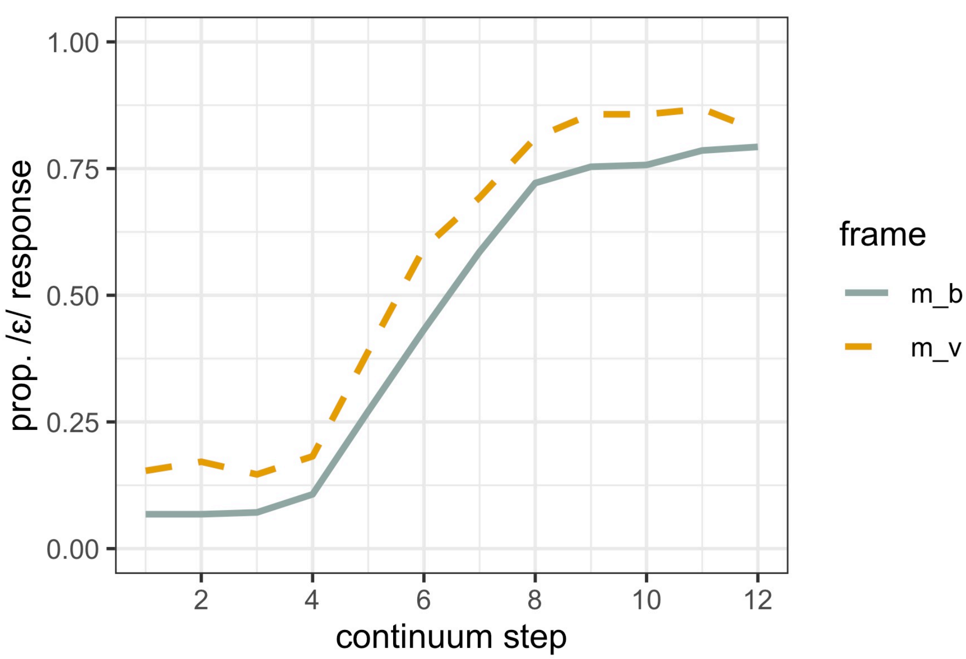

Figure 1: Experiment 1 categorization responses along the continuum (*x* axis, where step 1 is the most /æ/-like), split by consonant frame. The proportion of /ɛ/ responses is plotted on the *y* axis.

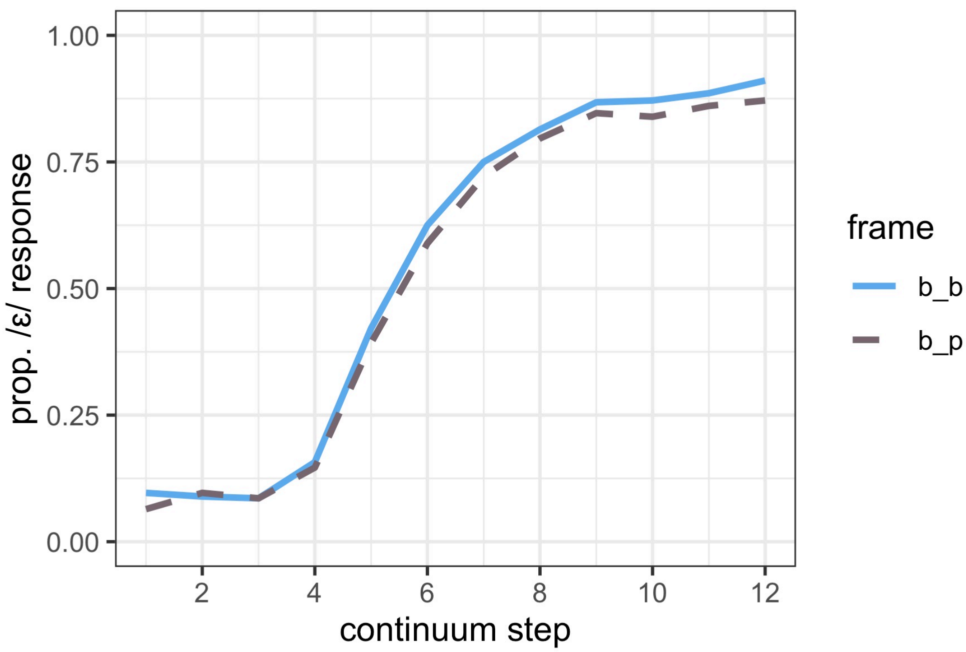

Figure 2: Experiment 2 categorization responses along the continuum, split by consonant frame.

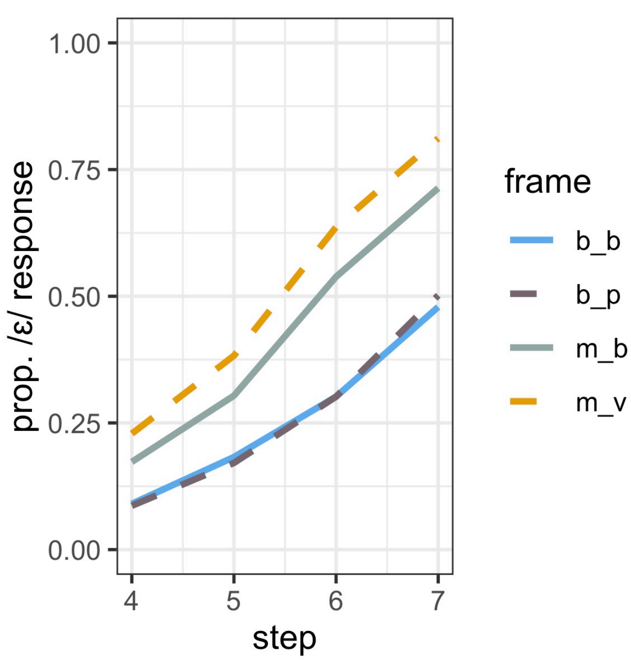

Figure 3: Experiment 3 categorization (click) responses along the continuum, split by consonant frame.

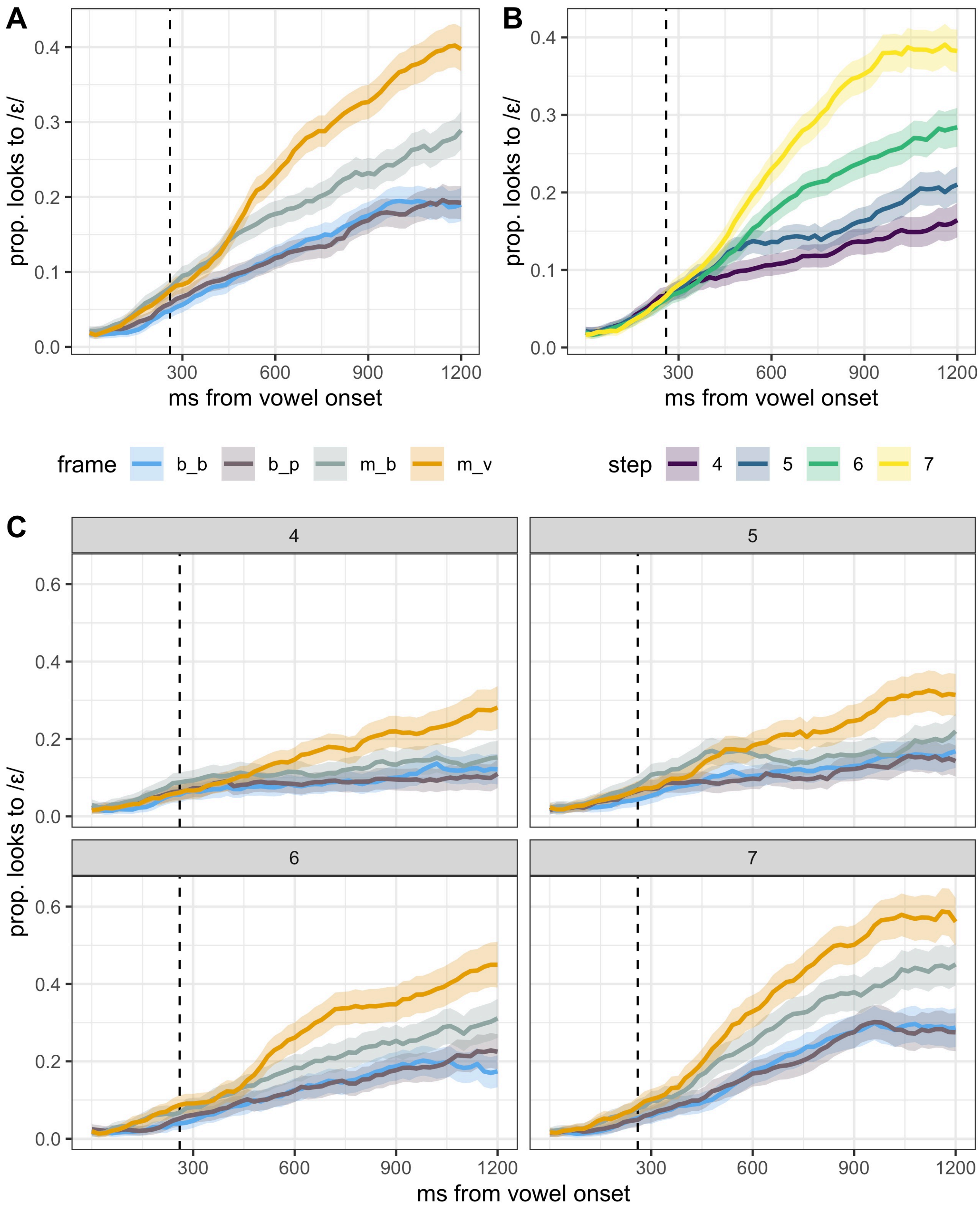

Figure 4: Experiment 3 eye movement data, split by consonant frame (panel A), by continuum step and frame (panel B), and by frame, split by continuum step (panel C). The proportion of looks to /ɛ/ over time are plotted, with 95% CI. The dashed vertical line indicates the vowel offset (260 ms).

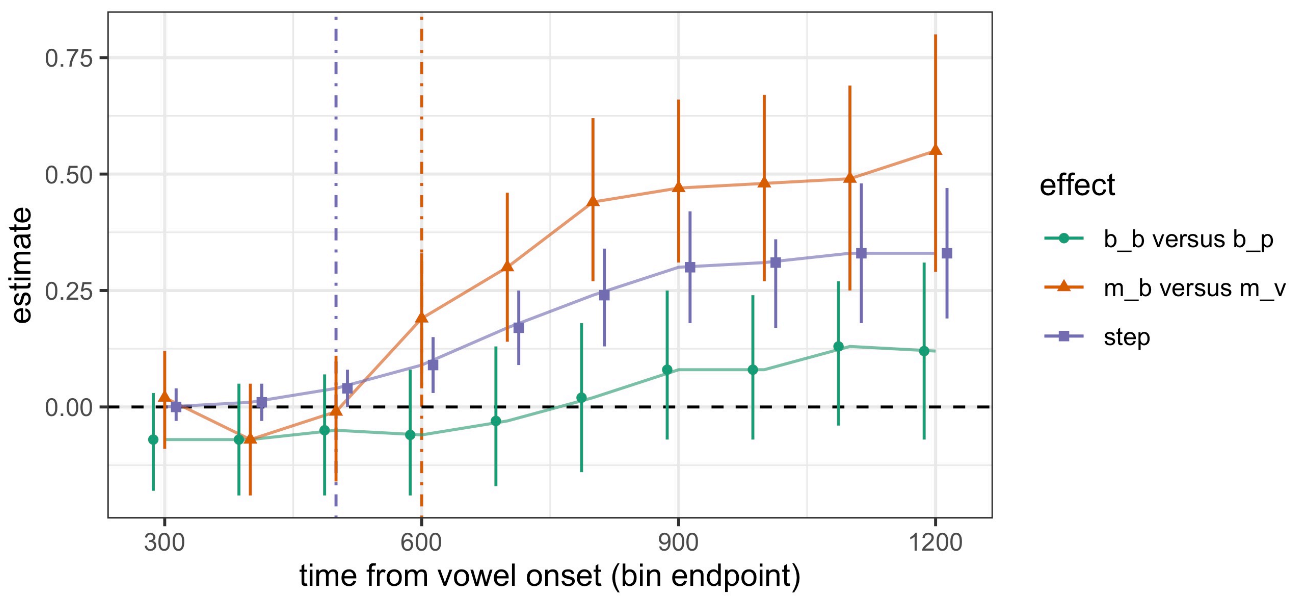

Figure 5: Model estimates for the moving window analysis, for continuum step, and two pairwise comparisons between frames of interest. The window starts at the time bin containing data for 200-300 ms from target onset, and proceeds in 100 ms intervals (300-400, 400-500, etc.). The dashed line is placed at 0; an estimate is reliable when its interval does not include 0, as indicated by vertical line showing the timing for continuum step, and for /mVb/ versus /mVv/ frames.

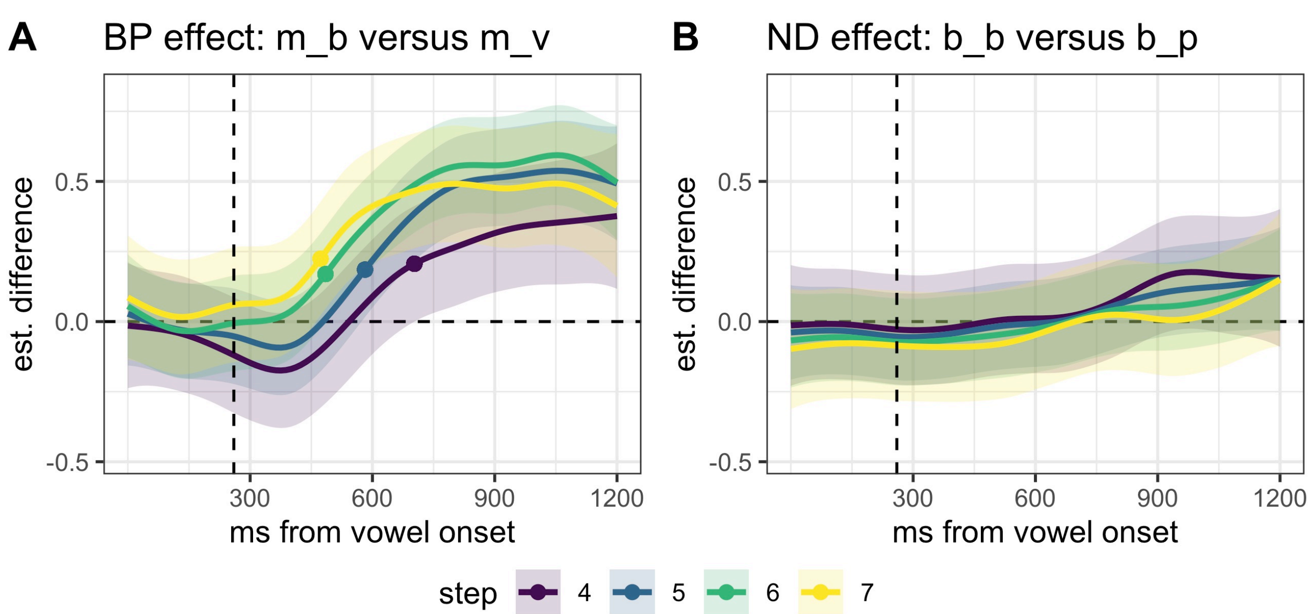

Figure 6: Difference smooths for consonant frame pairs (Panel A: /mVb/ versus /mVv/; Panel B: /bVb/ versus /bVp/). The point at each trajectory indicates when it has diverged from zero (see text). Step 4: 703 ms, Step 5: 582 ms, Step 6: 485 ms, Step 7: 473 ms). The dashed vertical line indicates the vowel offset (260 ms).

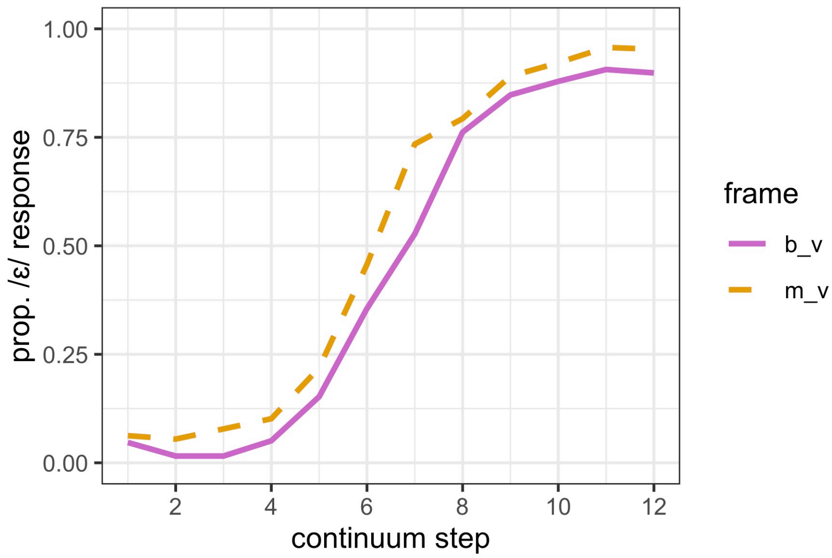

Figure 7: Experiment 4 categorization responses along the continuum, split by consonant frame.