Spoken word identification involves accessing position invariant phoneme representations

Jeffrey S. Bowers

Nina Kazanina

Nora Andermane

School of Experimental Psychology

University of Bristol

Key words: Phoneme; position specificity; spoken word identification, lexical access code.

Author Note:

Correspondence concerning this article should be addressed to Jeffrey Bowers, School of Psychology, University of Bristol, 12a Priory Road, Clifton, Bristol, BS8 1TU. e-mail: j.bowers@bristol.ac.uk. This research was funded by a Leverhulme Grant (RPG-205) awarded to the first author.

Abstract:

In two adaptation experiments we investigated the role of phonemes in speech perception. Participants repeatedly categorized an ambiguous test word that started with a blended /f/-/s/ fricative (*?ail* can be perceived as /fail/ or /sail/) or a blended /d/-/b/ stop (*?ump* can be perceived as /bump/ or /dump/) after exposure to a set of adaptor words. The adaptors all included unambiguous /f/ or /s/ fricatives, or alternatively, /d/ or /b/ stops. In Experiment 1 we manipulated the position of the adaptor phonemes so that they occurred at the start of the word (e.g., *farm*), at the start of the second syllable (e.g., *tofu*), or the end of the word (e.g., *leaf*). We found that adaptation effects occurred across positions: Participants were less likely to categorize the ambiguous test stimulus as if it contained the adapted phoneme. For example, after exposure to the adaptors *leaf*, *golf*... etc., participants were more likely to categorize the ambiguous test word *?ail* as ‘sail’. In Experiment 2 we also varied the voice of the speaker: Words with unambiguous final phoneme adaptors were spoken by a female while the ambiguous initial test phonemes were spoken by a male. Again robust adaptation effects occurred. Critically, in both experiments, similar adaptation effects were obtained for the fricatives and stops despite the fact that the acoustics of stops vary more as a function of position. We take these findings to support the claim that position independent phonemes play a role in spoken word identification.

Traditional linguistic theory postulates a small set of phonemes that can be sequenced in various ways in order to represent thousands of words in a language (Chomsky & Halle, 1968; Trubetzkoy, 1969). Phonemes are the smallest linguistic unit that can distinguish word meanings and usually are of a size of a single consonant or vowel, e.g., the consonants /b/ and /p/ are phonemes in English because they differentiate the words “bark” and “park”. Phonemes are critically distinguished from speech sounds (i.e. phones) in their level of abstractness. Phones are acoustically defined units that are often context-dependent, i.e. in a given language a certain phone may be bound to a specific syllable position, or require a certain stress pattern, or occur within the context of specific surrounding sounds. By contrast, phonemes are abstract entities that encompass several phones. For example, the phoneme /t/ is an abstract representational unit that in English is realized as an aspirated [t^h] syllable-initially as in *top*, as an unaspirated [t] following /s/ as in *star* or as an unreleased [t˺] in the syllable-final position as in *cat*. In other words, [t^h], [t] and [t˺] are different phones which in English represent a unique phoneme /t/.

A key theoretical reason for uniting distinct phones under the same phoneme category is that, despite their acoustic and articulatory differences, they operate as a single unit across a range of synchronic and historical language processes. Take the case of morphological derivation. Morphological derivation often leads to changes in the stress position that in turn result in differences in the quality of the vowel in the root morpheme. For example, the stressed vowel [ɒ] in *solid* [ˈsɒlɪd] changes to an unstressed [ə] in *solidity* [səˈlɪdɪti].[1] If phones were used to represent words, then there would be no *solid* in *solidity*. However, the existence of abstract phonemes ensures that *solidity* contains *solid* as the root morpheme. The same point can be

[1] Throughout the paper British English transcription will be used.

illustrated in pairs *compete* [k^{h}əmˈp^{h}iːt̚] – *competition* [ˌk^{h}ɒmpəˈt^{h}ɪʃən], *photograph* [ˈfəʊtəgrɑːf] – *photographer* [fəˈt^{h}ɒgrəfə^{r}] and indeed, is ubiquitous across the lexicon.

Another common effect of morphological derivation involves resyllabification of the final consonant of the root morpheme accompanied by a change in the acoustic identity of the consonant. For example, /t/ is realized as an unreleased [t̚] at the end of *float*, but as an aspirated [t^{h}] in *floatation*. This process is ubiquitous, e.g. *rate* [ˈreɪt̚] – *rated* [ˈreɪ.t^{h}ɪd], *type* [ˈt^{h}aɪp̚] – *typing* [ˈt^{h}aɪ.pɪŋ]. So once again phoneme representations are indispensable to preserve the compositionality of morphologically complex words.

In sum, the lexicon is much more regular – and perhaps easier to learn – if lexical representations are formulated in terms of phonemes rather than context-specific or position-specific phones. This may also explain why we employ a common written letter ‘t’ for the spelling of *top* and *cat* rather than one letter for [t^{h}] and another for [t̚].

Although phonemes are widely assumed in linguistic theory, the psychological evidence in support of phonemes, at least in the domain of speech perception, is scant. This has given rise to various models that abandon phonemes as a functional unit in speech perception. For example, on one view, words are stored and directly accessed by position-specific phones (or positional variants of phonemes in Pierrehumbert’s 2003 terminology). Pierrehumbert’s (2003) rationale for positional units (defined in terms of syllable or word position) stems from the observation that acoustic signature is more stable for position-specific phones compared to position-independent phonemes. These position-specific phones in turn map onto lexical representations.

Similarly, a number of computational models of spoken word identification (e.g., Luce, Goldinger, Auer, & Vitevitch, 2000; McClelland & Elman, 1986) bind segments to time in long-term memory in order to code for the order of segments. For example, in the TRACE model, different ‘d’ segments (d-at-time-1 and d-at-time-3) are used to activate *dog* and *god* representations, respectively. These time-bound segments can be seen as analogous to Pierrehumbert’s position-specific phones (in that the segments do not abstract across position) although the input units in these models are often labeled phonemes.

The common rejection of position invariant phonemes in psychological theories and models of word perception is a fundamental claim, and we explore this issue here. First we review the current empirical evidence regarding phonemes in the domains of speech production and perception, and then describe two experiments that provide strong evidence that phonemes do indeed play a role in word perception.

Empirical evidence for phonemes in speech production

In the domain of speech production the evidence for phonemes, i.e., segment-sized position-invariant units, is reasonably strong. One of the best pieces of evidence for segment-sized units comes from speech errors that involve swapping segments in corresponding syllable positions (e.g., swaps between onset consonants, such as “heft lemisphere” in lieu of “left hemisphere”). These swaps require positing segment size units (Fromkin, 1974). Evidence that the segment size units are coded independent of syllable position comes from swaps in non-corresponding syllable positions. For example, Vousden, Brown and Harley (2000) found that more than 20% of relevant phonological errors involved changes across syllable positions (e.g., *film* mispronounced as *flim*).

Priming studies point to a similar conclusion. For example, when participants are asked to name an object and its color, naming is facilitated by phoneme overlap between the color and object name both when overlapping segments occur in the same position (e.g. *green goat* vs. *red goat*) and when they occur in different syllable positions (e.g., *green flag* vs. *red flag*; Damian & Dumay, 2009). These findings lend support to the view that phonemes in speech production are coded independently of syllable position and are bound to syllable frames during production (e.g., Shattuck-Hufnagel, 1986).

Empirical evidence against phonemes in speech perception

Although phonemes are widely assumed in theories of speech production, it does not necessarily follow that phonemes are involved in speech perception as well. Indeed, Hickok (2014) recently developed a model of speech processing that holds phonemes as functional units in speech production but not perception. Consistent with this hypothesis, a number of psycholinguistic findings are taken to challenge the psychological reality of phonemes as units of perception, and this has led to a number of theories and models of speech perception that explicitly reject phonemes (e.g., Goldinger 1998; Luce et al., 2000; Oden & Massaro, 1978; Pierrehumbert, 2003). We review this data next.

Perhaps the most common experimental method used to challenge phonemes is perceptual learning. In these experiments participants learn to identify a degraded or distorted speech sound in one context, and the question is whether the learning generalizes to other contexts. It is assumed that generalization should extend to all allophonic forms of a given phoneme if indeed phonemes play a role in speech perception. By contrast, if generalization is restricted, it is taken as evidence against phonemes.

First consider a perceptual learning study in Dutch by Mitterer, Scharenborg, and McQueen (2013) in which no learning was observed between acoustically dissimilar allophones both within and between syllable positions. The phonemes /l/ and /r/ in Dutch each have at least two allophones: /l/ includes an alveolar lateral approximant [l] used in the syllable onset position ('light l'), and a velarized counterpart [ɫ] used in the syllable offset ('dark l'); /r/ includes an alveolar trill [r] and uvular trill [R] in onset position and, in addition, an alveolar approximant [ɹ] in the offset position. Mitterer et al. trained listeners to classify a novel morphed sound [ɫ/ɹ] (that was ambiguous between [ɫ] and [ɹ] in syllable offset position) as an /l/ or an /r/ by presenting it either in words that ended in /l/ (e.g. *acceptabel̲* 'acceptable') or in /r/ (e.g. *winter̲* 'winter'). An effect of training was found for new [ɫ/ɹ] morphs that occurred in the syllable-final position, but not for new morphs such as [ɫ/r] or [l/r] that included another allophone of /l/ and /r/ (regardless of position). On the basis of these findings the authors concluded that perceptual learning – and by extension speech perception – is mediated by allophones.[2]

Indeed, the results of number of perceptual learning studies have been taken as evidence that the relevant sublexical units are more acoustically specific than phonemes. For example, Dahan and Mead (2010) trained participants to identify consonants in noise-vocoded speech and found that generalization was greatly modulated by the degree to which training and test sounds were similar acoustically. That is, they found that consonants were easier to recognize when they occurred in the same syllabic position at training, when they were flanked by the same vowel, and

[2] Jesse and McQueen (2011) did show that learning to categorize a distorted fricative in a syllable-final position generalized to the perception of the same fricative in syllable-initial position, and they took this to support phonemes. However, fricatives are largely acoustically invariant across positions; hence the findings could be explained at the allophone level (Mitterer et al., 2013).

when spoken by the same speaker. Based on these results, the authors hypothesized that the sub-lexical perceptual categories that support speech perception are even more specific than allophones. Consistent with this conclusion, a number of authors have found that perceptual learning is often voice specific (e.g., Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2007).

A similar conclusion was reached by Reinisch, Wozny, Mitterer and Holt (2014). These authors found that listeners who learnt to categorize an ambiguous [b/d] sound in the context of the vowel /a/ (i.e. *a_a*) as /b/ or /d/ during a learning phase did not generalize to the *u_u* context even though an acoustic encoding of the /b/ vs /d/ distinction is similar in the two vowel contexts. Based on the context specificity of their learning effects the authors concluded:

> "From a theoretical perspective the results of the present study suggest that pre-lexical processing does not make use of abstract phonological features, context-free phonemes, or speech gestures." (p.104, Reinisch et al., 2014)

Even more dramatically, perceptual learning is sometimes ear specific, highlighting that it can be closely tied to the sensory specifics of the learning context rather than to abstract categories such as phonemes (Keetels, Pecoraro, & Vroomen, 2015). The fact that generalisation is restricted and context-sensitive in the above perceptual learning studies is taken as evidence against phonemes.

In addition to these perceptual learning studies, findings from long-term priming studies are often taken as evidence that word identification is mediated by perceptually specific as opposed to abstract phoneme representations (e.g., Goldinger, 1996). For example, Pufahl & Samuel (2014) found that priming was greater when words were repeated in the context of the same environmental sounds at study and test (e.g., a phone ringing). The authors took this to suggest that non-linguistic sounds are part of the stored phonological representations that support word identification.

At the same time that some authors reject phonemes in favour of smaller and more detailed sub-lexical categories such as allophones (Mitterer et al., 2013), others reject phonemes in favour of larger sub-lexical representations, including demi-syllables and syllables. For example, consider some classic selective adaptation studies. Ades (1974) found that multiple repetitions of a syllable starting with the consonant *d* (e.g., [dæ]) led to a shift in the categorical boundary in the /dæ/-/bæ/ continuum towards /bæ/ (i.e., participants responded /bae/ more often). Critically, however, repeated presentation of the syllable [æd] that contained the consonant *d* in the final position did not affect the perception of the /dæ/-/bæ/ continuum. This suggests that the initial and final *d*'s are coded separately. Similarly, Samuel (1989) found that multiple repetitions of a syllable such as [ba] led to a shift in the categorical boundary in /ba/-/pa/ continuum towards /pa/, but failed to affect perception of the same consonants in the syllable-final /ab/-/ap/ continuum. Based on these findings Samuel rejected position independent phonemes in favor of demi-syllables as access codes to words (also see Fujimura, 1976; Rosenberg, Rabiner, Wilpon, & Kahn, 1983).

Massaro (1975) argued that syllables rather than phonemes act as sub-lexical perceptual representations given that many consonant phonemes cannot be perceived in isolation (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). For example, the consonant *d* cannot be recognized separately from its coarticulated vowel. That is, when the duration of the vowel in the consonant-vowel syllable (CV) is gradually decreased, there is no point in which the perceiver just hears the *d*. Instead, the CV syllable is perceived as a complete syllable until the vowel is eliminated almost entirely. Once further signal is removed, a nonspeech whistle is heard. The conclusion is that the CV syllable is perceived as an indivisible whole or gestalt, and not decomposed into phonemes. In addition, Massaro (1975) rejects phonemes based on the claim that there are no invariant acoustic features for many consonant phonemes that would allow them to be identified separately from the following vowel (for a critical review of the phoneme, see Goldinger & Azuma, 2003).

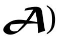
In Defense of Phonemes:

Despite aforementioned findings, there are a number of reasons why we think it is premature to reject phonemes in speech perception. First, and perhaps most importantly, all the above criticisms do not grapple with the strong linguistic arguments in support of phonemes. For example, as noted above, phonemes often preserve the compositionality of morphologically complex words (e.g., so that there is *solid in solidity* and *float* in *floatation*). If phonemes are to be rejected, then some account of the morphological structure of words needs to be offered.

Second, much of the evidence presented against phonemes rests on theoretical confusions. Perhaps most importantly, no one claims that phonemes are <u>the</u> sub-lexical unit of perception (that is, the only sub-lexical unit). Rather, the claim is that phonemes are <u>a</u> sub-lexical unit of perception (that is, one of perhaps several sub-lexical units). Accordingly, perceptual learning studies that provide evidence for allophones or other sub-lexical units do not challenge the existence of phonemes. They just show that there are other sub-lexical representations involved in speech perception as well. Indeed, given that perceptual learning can be restricted to one ear

(Keetels et al., 2015), it seems that these tasks are sensitive to low-level perceptual processes that fall outside the domain of speech perception altogether.

Furthermore, upon closer inspection, the logic of some of these perceptual learning studies is unclear. Consider again the Mitterer et al. (2013) study. Listeners were trained to identify an ambiguous [ɫ/ɹ] morph (produced by combining dark [ɫ] allophone of /l/ with approximant [ɹ] allophone of /r/) in syllable-final position as an /l/ or /r/, and then at test categorized novel instances of [ɫ/ɹ] morph in syllable-final position as /l/ or /r/. That is, the perceptual space of the allophones [ɫ] or [ɹ] was altered to incorporate acoustically similar inputs (various blends of [ɫ/ɹ]). This shows that learning took place at the allophonic level. The key finding taken as inconsistent with phonemes is that this learning did not impact on the perception of acoustically dissimilar allophones of /l/ and /r/. For instance, after training with the [ɫ/ɹ] morph, the perceptual space associated with the light [l] allophone remained unaltered and did not incorporate the morph [l/r].

There is a problem with this line of reasoning, however. Namely, no theory should expect any generalization given that the [l/r] morph is novel and acoustically distinct from the [ɫ/ɹ] morph (given that [ɹ] and [r] are acoustically dissimilar). To illustrate, consider an analogy with written letters. It is widely agreed that visual word identification involves accessing abstract letter codes that map together different visual forms of letters, even unrelated visual forms (Bowers, Vigliocco, & Haan, 1998; Coltheart, 1981; McClelland, 1976). For example, it is widely assumed that different exemplars of lower-case *a* and upper-case *A* map onto the abstract code *A**. Nevertheless, if a participant learns that that a new distorted version of upper-case *A* (e.g., ) is a member of the category *A**, it provides no basis for the reader to reorganize his or her perceptual space of lower-case *a*. Accordingly, the failure to modify the perceptual space of lower-case *a* provides no reason to reject abstract letter identities. The same is true for the sounds in the Mitterer et al.'s (2013) study.

The logic of some other studies used to support syllables rather than phonemes is also unclear. For example, consider the finding that some consonant phonemes are not accessible to phenomenological experience (e.g., you cannot perceive a /b/ in isolation). Although Massaro (1975) uses this to argue against phonemes, this seems a weak basis for rejecting phonemes: there are presumably many representations that are not accessible to introspection. Similarly, the claim that there are no invariant acoustics associated with some phonemes provides no basis for rejecting phonemes. Consider again the analogous case of visual letters. No one would conclude that abstract letter codes do not exist simply because there is no visual invariance between an A/a. Rather, the relevant question is how abstract letter codes might be learnt (e.g., see Bowers & Michita, 1998). In the same way, the fact that there is no obvious acoustic invariance between stop consonants in different contexts does not rule out phonemes. It just means that the listener would have to learn how to map acoustically distinct allophones onto abstract phonemes.

A third general reason we think it is premature to reject phonemes in perception is that phonemes are well accepted in the domain of speech production (e.g., Dell, 2014; Fromkin, 1974; Hickock, 2014). At the very least, evidence for phonemes in production increase the *a priori* plausibility that similar (or the same) representations are involved in perception as well.

Finally, a few empirical studies do provide some evidence that phonemes play a role in perception. For example, Morais, Castro, Scliar-Cabral, Kolinsky and Content (1987) presented pairs of CVCV words dichotically to literate and illiterate Portuguese speakers and asked them to identify the word presented to either the left or right ear (the target ear varied across trials). A common error for both groups involved single segments, such as initial consonant migrations, with the initial consonant of the to-be-ignored words migrating to the corresponding position in the target word. This provides evidence for segment-sized units of perception. The fact that illiterates showed the same pattern of results demonstrates that these units are not a by-product of learning a written alphabet. Similarly, Cutting and Day (1975) reported phonological fusions in dichotic listening tasks in English in which segments in the to-be-ignored input were added to the segments in the target input (rather than replacing a segment in the same position). For example, the presentation of *banket*/*lanket* led to the identification of *blanket*. As argued by Morais, Castro, Scliar-Cabral, Kolinsky and Content (1987), if syllables rather than phonemes were the smallest unit of perception, it is not clear why two CVC inputs (*ban* and *lan*) would result in the perception of a CCVC syllable *blan* (rather than combine into a CVCCVC string, for example).

More recently, Toscano, Anderson, and McMurray (2013) provided evidence that phonemes are coded independently of position. They found that anadromes – words that share the same phonemes but in the opposite order, such as *sick* and *kiss* – are confusable. In a visual world paradigm study, participants looked at foil anadrome pictures (e.g., cat) in response to a spoken target word (e.g., *tack*) more than control items. Note that the effect cannot be explained at the allophone level as there is an aspirated [k^h] in *cat* and an unreleased [k˺] in *tack* (similarly, ‘t’ is aspirated in *tack* but unreleased in *cat*), and points to the existence of phonemes /k/ and /t/ that are independent of syllable position.

We readily acknowledge, however, that the psycholinguistic evidence in support of phonemes in perception is mixed, and more evidence is needed before any strong conclusions are warranted. In the current paper we investigate the role of phonemes in speech perception by using an adaptation task in which participants categorized test stimuli that can be perceived as one word or another due to an ambiguous initial phoneme (e.g., *?ail* can be perceived as *fail* or *sail*; on repetition of ?*ail* the perception oscillates between the *fail* and *sail*, much like a Necker cube in vision). We assess how participants categorized the test stimuli as a function of preceding adaptor words that contained the critical unambiguous phoneme (e.g., /f/ or /s/). In Experiment 1 the critical unambiguous phoneme occurred at the beginning of the adaptor word (initial position; e.g., *fact*), in the onset of the second syllable (medial position; e.g., *profit*), or in the coda of the final syllable (final position; e.g., *beef*). We manipulated the extent to which the acoustics of the critical phoneme within the adaptor words varies across positions by including fricative adaptors (/f/ and /s/) that do not vary greatly across position, as well as stop consonant adaptors (/d/ and /b/) that vary much more across positions. In Experiment 2 we assessed adaptation when the adaptor words containing the critical unambiguous phoneme in final position were pronounced by a female and the test words containing the ambiguous phoneme in the initial position were pronounced by a male. Robust adaptation effects across position and voice in Experiments 1 and 2 would provide evidence for abstract, position-independent phoneme representations. It is important to emphasize that adaptation in this task is manifest as a *reduced* likelihood of identifying the test stimulus as the word starting with the adaptor phoneme. For example, adaptation is found when a participant is less likely to identify the target as *fail* after being exposed to adaptors containing the /f/ phoneme. Accordingly, any adaptation effects cannot be attributed to a strategy of categorizing the targets to correspond with the adaptors.

Experiment 1

Method

Participants

Ninety-six adult participants took part in this study. Participants were undergraduate psychology students at the University of Bristol and adults from the Bristol community. All participants were native English speakers with no history of dyslexia and normal or corrected-to normal vision.

Stimuli

We recorded an adult male native English speaker (British English, Received Pronunciation) using a SHURE SM48 vocal microphone and Cool Edit Pro. Two types of stimuli were recorded, namely, adaptor words and word pairs used for creating test stimuli. The test word pairs differed in their first phoneme (and first letter) and started with fricatives (/f/ and /s/, e.g., *funny-sunny*) and stop consonants (/b/ and /d/, e.g., *beer-deer*). Thus, the critical ambiguous phoneme in the test stimulus was always syllable- and word-initial. We focused on fricatives and stop consonants because the acoustics of fricatives are relatively invariant across position whereas stop consonants vary a great deal.[3] We started with 16 test word pairs (8

[3] In both syllable-initial and final positions, /s/ and /f/ are characterised by a salient and relatively long-lasting friction noise. /s/ is characterised by a clear distinct spectral shape and exhibits a primary spectral peak usually between 4-7 kHz independent of syllable position, whereas the spectrum for /f/ is usually flat (Reetz & Jongman, 2009). Acoustic measurements confirmed that this held for the fricatives in our adaptor stimuli. Furthermore, the amplitude of the friction noise was higher for /s/ than /f/ (by 16 and 11 dB in the initial and final positions respectively), which accords with previous reports (e.g., Reetz & Jongman, 2009). The duration of the friction noise was longer for /s/ than for /f/ in the initial position (283 vs 256 ms) but similar in the final position (both 320 ms).

Acoustic correlates of stops, on the other hand, differ significantly depending on their syllable position. Syllable-initially, the primary correlates are the burst quality (if the fricative pairs, 8 stop consonant pairs) and blended them using Tandem-STRAIGHT speech modification software resulting in a morphing continuum of 30 steps that perceptually changes from one word to the other (e.g., continuum starting with the word *fail* and ending with the word *sail*). We chose 3-5 blends from the midpoint of each continuum for a pilot study. Eighteen participants decided whether each blend started with /f/ or /s/, or /b/ or /d/ and we chose the most ambiguous blends as the test words; these were the words *?ail* (from *fail* & *sail*), which was categorised as *fail* 58%, and as *sail* 42 % of the trials, and the blend *?ump*, which was categorised as *bump* 45%, and as *dump* 55% of the trials.

Adaptor words were chosen by selecting words with the critical phoneme (e.g., /d/, /b/, /s/ and /f/) in the initial (e.g., *fly*), the medial (e.g., *tofu*), or the final (e.g. *leaf*) position. Seventy-five adaptor words were selected for the /d/, /b/, /s/, and /f/ conditions; 25 words per position. The duration of the adaptor words ranged from approximately 350 to 900 ms. No adaptor word contained the opposing critical phoneme (e.g. adaptor words starting with /f/ did not contain /s/ anywhere within the word). The average duration of the adaptor words was 764 ms in the fricative condition and 571 ms in the stop condition. The full list of adaptor words can be found in Appendix 1.[4]

Design and Procedure

burst is present), voice onset time (VOT) and formant transitions into the following vowel (Johnson, 2011; Reetz & Jongman, 2009). In our sample a discernible burst was present in in 22/25 of /b/-initial adaptors and in all (25/25) /d/-initial adaptors; the VOT values were 8.4 and 6.1 ms for initial /b/ and /d/ respectively (and typical for English voiced consonants). Word finally, two of the acoustic features as defined above – VOT and formant transitions – are non-existent due to the lack of a following vowel. The release was also weaker and present in 11/25 and 25/25 of word-final /b/ and /d/ adaptors respectively. The most reliable acoustic correlate of a word-final stop is formant transitions in the preceding vowel. In sum, there is little invariance in acoustic realisation of word-initial vs. final stop consonants.

[4] One word in the final-/s/ adaptor condition, *lens*, was included incorrectly (ends in [z] rather than [s]).

Each participant was assigned to one of the initial, medial, or final adaptor conditions (32 participants in each). The experiment was divided into four blocks, each corresponding to one of the four adaptor phoneme conditions (/f/, /s/, /b/, & /d/), with the order of block presentation counterbalanced. Each block contained six adaptation phases and associated test phases; the initial adaptation phase was approximately 3 minutes long, and the remaining five adaptation phases were about 1 minute long. During the initial adaptation phase participants passively listened to 25 adaptor words repeated randomly 9 times each in /f/ and /s/ adaptor conditions and 12 times each in /b/ and /d/ adaptor conditions (in order to equate the total time of the adaption phase). In the following adaptation phases the items were repeated 3 and 4 times each, respectively. After each adaptation phase, participants were presented with ten test trials and asked to categorise the ambiguous test stimulus (e.g., categorise *?ail* as *fail* or *sail*). All stimuli were delivered via headphones. Each experimental session consisted of 240 test trials and took approximately 60 minutes to complete.

During each adaptation phase, participants passively listened to adaptor words while an asterisk was visible in the centre of the screen. Following this, a beep was presented for 1000 ms, followed by a 3000 ms pause, and then the 10 test trials. On each test trial the response cue "f ←   →s" or "b←   →d" was presented for 3000 ms in the centre of the screen, and simultaneously, the auditory test stimulus (*?ail* or *?ump*) was presented. Participants were asked to categorise the test stimulus by pressing the left or right Shift keys on the keyboard (left for 'fail' and 'bump', right for 'sail' and 'dump'). Participants responded while the response cue was on screen. After a response, the screen was cleared, there was an inter-trial interval of 500 ms.

After completing the ten test trials, participants took a short break before commencing the next adaptation phase.

We arbitrarily assigned ‘sail’ and ‘dump’ as target responses for the stop and fricative conditions, respectively. Accordingly, the /s/-adaptors are considered as target congruent in the fricative condition and the /d/-adaptors are considered target congruent in the stop condition. Adaptation effects reflect the impact of the adaptor words on the categorisation of the ambiguous test stimulus, with a reduction in the number of target congruent categorizations. For example, a reduction in categorising the ambiguous test stimulus *?ail* as ‘sail’ following /s/-adaptor words compared to /f/-adaptor words.

Results

The data points excluded from the analysis were response times of 3000 ms or greater, as well as response times shorter than 200 ms. Overall, 1.6% of responses were excluded based on these criteria; 2.0 % from the initial position condition, 1.2% from the middle position condition, and 1.5% from the final position condition. The mean percentages of target ‘sail’ and ‘dump’ responses were computed for each participant in each adaptation condition, and the results can be seen in Figure 1.

The data from Experiment 1 were entered into a 2×2×3 mixed samples ANOVA with consonant type (fricative vs. stop) and adaptation condition (target-congruent vs. target-incongruent) as within-subject factors and position (initial vs. medial vs. final) as a between-subject factor. There was a marginally significant effect of consonant type (F (1, 95) = 3.46, p = .066, ηp2 = .04) and a significant effect of adaptation condition (F (1, 95) = 110.04, p < .001, ηp2 = .54). The interaction between adaptation condition and position was significant (F (2, 94) = 18.92, p < .001, ηp2 = .29), whereas the interaction between consonant type and position was not

(*F* (2, 94) = .61, *p* = .55, ηp2 = .01). Finally, the three-way interaction was not significant (*F* (2, 94) = .82, *p* = .44, ηp2 = .02), suggesting similar adaptation effects across positions for fricatives and stops.

Next fricatives and stops were considered separately. First consider the adaptation effects for the fricatives across position. We carried out a 2×3 ANOVA that included adaptation condition (target-congruent vs. target-incongruent) as a within-subject factor and position within the adaptor word (initial vs. medial vs. final) as a between-subject factor. The analysis showed a large main effect of adaptation condition such that the ambiguous test stimulus was more likely to be categorized as 'sail' having been adapted with /f/-adaptors and as 'fail' following /s/-adaptors (*F* (1, 95) = 81.12, *p* < .001, ηp2 = .47). The interaction between adaptor condition and position was significant (*F* (2, 94) = 10.94, *p* < .001, ηp2 = .19) and largely driven by the fact that the adaptation effect was largest with adaptor words containing the critical phoneme in the initial position. However, strong adaptation effects were found in all positions. A set of planned contrasts showed that the overall adaption effect was significant in the initial (*t* (31) = 6.80, *p* < .001, *d* = 1.45, two-tailed), medial (*t* (31) = 4.47, *d* = .52, *p* < .001, two-tailed), and final (*t* (31) = 3.87, *p* = .001, *d* = .41, two-tailed) position conditions.

A similar outcome was obtained for the stop consonants. A parallel 2×3 ANOVA showed a large effect of adaptor condition such that the ambiguous test stimulus was more likely to be categorized as 'dump' after having been adapted with /b/-adaptors and 'bump' following /d/-adaptors (*F* (1, 95) = 38.07, *p* < .001, ηp2 = .29), and an interaction between adaptor condition and position (*F* (2, 94) = 9.96, *p* < .001, ηp2 = .18), largely driven by the fact that the adaptation effect was largest with adaptor words starting with the critical phoneme. One difference, however, was that the adaptation effects were not found in all three positions. A set of planned contrasts showed that the adaption effect was significant in the initial (*t* (31) = 5.32, *p* < .001, *d* = .96, two-tailed) and final (*t* (31) = 2.83, *p* = .008, *d* = .25, two-tailed) positions. However, this difference (4.38%) did not achieve significance in the medial position, (*t* (31) = 1.66, *p* = .108, d = .13, two-tailed).

We observed robust adaptation effects in five out of six conditions; what should be made of the failure to obtain significant adaptation effect for our ambiguous stop target (*?ump*) with medial adaptors? One possibility is that this reflects a methodological weakness of the study, namely, a subset of participants may have consistently heard ‘bump’ or ‘dump’ regardless of the adaptors (that is, the test stimuli may not have been ambiguous for everyone). Such ceiling or floor effects would reduce overall adaptation effects. In order to explore this possibility we carried out a second set of analyses in which we excluded all responses from those participants who reported /s/ or /d/ on over 90% of trials or under 10% of the trails (if a participant showed a ceiling/floor effect for fricatives but not for stops we kept the stop responses and vice versa). Based on this criterion we excluded the following data: For fricatives: 9 out of 32 participants were excluded from the initial position condition, 10 out of 32 from the middle position condition, and 14 out of 32 from the final position; For stops: 14 out of 32 participants were excluded from the initial position, 18 out of 32 from the middle position, and 19 out of 32 from the final position. The adaptation effects for the remaining data are shown in Figure 2.

Insert Figure 2 about here

The overall pattern of findings was similar to above, with one critical difference, namely, that adaptor effects were now significant for fricatives and stops in all conditions. For fricatives, a set of planned contrasts showed that the adaption effect was significant in the initial (*t* (22) = 9.89, *p* < .001, *d* = 3.04, two-tailed), medial (*t* (21) = 5.58, *p* < .001, *d* = 1.07, two-tailed), and final (*t* (17) = 4.78, *p* < .001, *d* = 1.14, two-tailed) position conditions. Critically, the same was found with stops, with significant adaptation effects obtained in the initial (*t* (17) = 8.95, *p* < .001, *d* = 2.00, two-tailed), medial (*t* (13) = 2.41, *p* = .031, *d* = .59, two-tailed), and final (*t* (12) = 3.43, *p* = .005, *d* = .90, two-tailed) positions.

Discussion

The results of Experiment 1 provide good evidence for position independent phonemes. The categorization of ambiguous fricatives in the syllable-onset position in the test words was altered by adaptor words with the critical phoneme in the word-initial, medial or final position. Similar results were obtained for ambiguous stops following word-initial and word-final adaptors. Once floor and ceiling effects were removed, adaptation effects occurred for both fricatives and stops across all positions.

The most critical finding is that adaptation effects were found with stop adaptors in the final position despite substantial differences in the acoustic realisation of stops across syllable position. In fact, the overall size of the adaptation effects for the final fricatives (14.1%) and stops (9.6%) was not so different. When final position adaptation effects are assessed in relation to the size of the effect in the initial position (28% for fricatives and 42.5% for stops), then fricatives (9.6/28 = 34%) and stops (14.1/42.5= 33%) showed the same degree of abstraction. It is also interesting to highlight the adaptation effects observed in the medial stop and fricative conditions given that in the majority of adaptors the second syllable was unstressed. Hence, the critical phoneme often occurred in an acoustically less salient, unstressed syllable in the adaptors but in an acoustically salient stressed syllable in the test stimulus. Indeed, 48 out of 50 stop adaptor words were stressed on first syllable (e.g., *robin,* *idol*; the only second-syllable stressed words were *Tibet* and *undo*), and 35 of 50 fricative adaptors were stress-initial (such as *awful*, *gossip*). This again shows that adaptation generalized over acoustic changes.

Experiment 2

The most critical result from Experiment 1 is that the adaptation effect generalized from word-final to word-initial stops even though acoustic realisation of stops varies quite dramatically as a function of syllable position. Accordingly, we thought it was important to replicate this finding, particularly so given previous failures to obtain adaptation across positions with stops in a similar procedure (Ades, 1974; Samuel, 1989). In Experiment 2 we assessed adaptation from final adaptors (both fricatives and stops), and in addition, we varied the speaker. That is, we included the same ambiguous target stimuli (spoken by a male) and newly recorded adaptors (spoken by a female). Accordingly, acoustic differences between the critical phonemes in the adaptors and test stimuli varied even more substantially than in Experiment 1.

Participants

Forty-eight adult participants took part in this study. Participants were undergraduate psychology students at the University of Bristol and adults from the Bristol community. All participants were native English speakers with no history of dyslexia and normal or corrected-to normal vision.

Stimuli

The ambiguous test stimuli *?ail* and *?ump* from Experiment 1 were again used as test stimuli in Experiment 2. New adaptor words were recorded for Experiment 2 using the same procedure as in Experiment 1; however the adaptor words were recorded with an adult female native English speaker (British English, Received Pronunciation) to increase the acoustic dissimilarity between adaptor words and targets. Overall 60 monosyllabic adaptor words were recorded; 15 of each contained the critical phonemes (/f/, /s/, /b/, and /d/) in the final position of the word. The duration of the adaptor words ranged from approximately 500 to 900 ms. The average duration of adaptor words was 768 ms for fricatives and 710 ms for stop consonants. The full list of adaptors is given in Appendix 2.

Design and Procedure

We followed the same procedure as in Experiment 1. The experiment was divided into four blocks, each corresponding to one of the adaptor conditions (/f/, /s/, /b/& /d/) in the final position; the order of block presentation was counterbalanced. Each block contained six adaptation phases and associated test phases; the initial adaptation phase was approximately 3 minutes long, and the remaining five adaptation phases were approximately 1 minute long. During the initial adaptation phase participants passively listened to 15 adaptor words repeated 16 times each. In the following adaptation phases the adaptor words were repeated 5 times each. After each adaptation phase, participants were presented with ten test trials and asked to categorize auditory test stimuli (i.e., categorize ‘?ail’ as *fail* or *sail* or categorize ‘?ump’ as *bump* or *dump*). Each experimental session consisted of 240 test trials and took approximately 60 minutes to complete.

Results

The data points excluded from the analysis were response times of 3000 ms or greater, as well as response times shorter than 200 ms; 1.4% of responses were excluded from analysis based on these criteria. As in Experiment 1, we assigned ‘sail’ and ‘dump’ as target responses for the stop and fricative conditions, respectively. The mean percentages of target ‘sail’ and ‘dump’ responses were computed for each participant in each adaptation condition, and the results can be seen in Figure 3. A 2×2 repeated samples ANOVA with factors consonant type (fricative vs. stop) and adaptation condition (target-congruent vs. target-incongruent) was applied to the data from Experiment 2. There was a significant main effect of adaptation condition (*F* (1, 47) = 21.88, *p*< .001, ηp2 = .32), but no significant effect of consonant type (*F* (1, 47) = 0.08, *p* = .780, ηp2 = .002), and no interaction between consonant type and adaptation condition (*F* (1, 47) = 1.01, *p* = .320, ηp2 = .02). Planned pairwise comparisons revealed significant adaptation effects for fricatives (*t* (47) = 2.89, *p*= .006, *d* = .23, two-tailed), reflecting the higher proportion of ‘sail’ responses after listening to /f/ adaptors and ‘fail’ responses following /s/ adaptors, and for stops (*t* (47) = 3.42, *p*= .001, *d* = .34, two-tailed) reflecting the higher proportion of ‘dump’ responses following /b/ adaptors and ‘bump’ responses following /d/ adaptors.

Insert Figure 3 about here

As in Experiment 1 we also assessed the adaptation effects when floor and ceiling effects were removed. That is, we removed all of the responses of those participants who responded /s/ or /d/ over 90% or under 10% of the time. For the fricative analysis 24 out of 48 participants were excluded, and for the stop analysis 28 out of 48 participants were excluded. Now the percentage of /s/ responses in the fricative condition was 58.3% after exposure to /f/ adaptors and 43.1% following /s/ adaptors (adaptation effect of 15.2%). In the stop consonant condition, the percentage of /d/ responses was 65.8% following /b/ adaptors and 35.0% following /d/ adaptors (adaptation effect of 30.8%). Pairwise comparisons showed a significant adaptation effect for fricatives (*t* (23) = 2.77, *p* = .011, *d* = 0.72, two-tailed), as well as for stop consonants (*t* (19) = 3.82, *p* = .001, *d* = 1.15, two-tailed).

Discussion

The key finding is that we obtained robust adaptation effects between syllable-final and syllable-initial stops and fricatives despite the fact that the adaptor and test stimuli were recorded by a female and male speaker, respectively. If anything, the adaptation effects were larger for the stops compared to the fricatives. Again, we take these findings to support the conclusion that phonemes play a role in speech perception.

General Discussion

Our key finding is that robust adaptation effects occurred between phonemes in different word positions (e.g., adaptor: *leaf*; test: *?ail*). Critically, the magnitude of the adaptation effect across positions was similar for fricatives that are relatively invariant in their acoustic form across positions and for stops that are less invariant across positions. Indeed, in Experiment 2, adaptation effects were obtained between syllable-final stop adaptors spoken by a female and syllable-initial stops spoken by a male. We take these findings to support the claim that position independent phoneme representations play a role in speech perception (e.g., Cutler, 2008; Toscano et al., 2013), and to challenge theories that reject phonemes in language perception (e.g., Hickock, 2014; Pierrehumbert, 2003).

In Experiment 1 we did observe that adaptation effects were largest when the critical phoneme was in the same position in the adaptor and test stimulus (e.g., adaptor: *farm*; test: *?ail*). Indeed, adaptation effects across positions were only ~1/3 the size as within position. However, the reduced adaptation effects across position do not necessarily undermine our conclusion that phonemes are coded independent of their position. It is well known that adaptation effects can occur at multiple levels of processing (Samuel & Newport, 1979) and are sensitive to both acoustic and phonological dimensions. Accordingly, it is likely that our adaptation effects also reflected a combination of lower- and higher level representations, with low-level acoustic representations playing more of a role in initial position condition (given the greater acoustic overlap here). The critical point is that robust adaption effects occurred across word and syllable positions, and these effects cannot be reduced to the acoustic similarity of phonemes across positions given the similar results were obtained with stops and fricatives.

In contrast, our findings are difficult to reconcile with theories that reject position-invariant phoneme representations. For example, if the sub-lexical units of speech identification are positional variants of phonemes, as proposed for example by Pierrehumbert (2003), then it is not clear why such robust adaption effects were obtained across syllable and word positions. Similarly, the hypothesis that the sub-lexical units of speech perception are syllables (e.g. Oden & Massaro, 1978) or demi-syllables (e.g., Samuel, 1989) is difficult to reconcile with our findings. On the syllable view, word-final adaptors such as *food* or *kid* activate the syllable [fud] or [kɪd] which are not organised into phonemes. Hence it is unclear why [fud] or [kɪd] decreases the likelihood of perceiving the ambiguous stimulus [?ump] as [dump], especially given that there is little acoustic similarity between the final and initial consonant *d*. Similarly, on the demi-syllable view, word final adaptors such as *cord, need, pond* activate demisyllables [ɔːd], [iːd], [ɒnd] which bear little overlap with the demisyllable [dʌ] of the target *dump.* In addition, syllables and demi-syllables do not account for the theoretical linguistic observations regarding the compositionality of morphologically complex words. For example, demi-syllables mask the relationship between members of a morphological family: *kiss* represented as a combination of demisyllables [kɪ] and [ɪs] is not present within *kissing* that is composed of [kɪ], [sɪ] and [ɪŋ].

How can we reconcile our findings with previous psycholinguistic research that has failed to observe evidence for phonemes? As detailed in the Introduction, the rejection of phonemes based on previous results often reflects a theoretical confusion. For instance, the fact that perceptual learning is often highly perceptually specific (in some cases ear specific; Keetels et al., 2015) says nothing about whether the perception also includes abstract phoneme representations. But we would note that there is one set of findings that does seem at odds with our results; namely, the previous adaptation studies that failed to obtain adaptation across syllable positions in non-lexical targets (Ades, 1974; Samuel, 1989). Why the difference?

Our speculative explanation for the contrasting results is as follows. First, our adaptation procedure was developed in response to our observation that the ambiguous test words regularly swapped with repetition (from 'dump' do 'bump' for example). This striking illusion suggested to us that these stimuli might make a sensitive measure of adaption, especially in light of their lexical status. As noted earlier, adaptation effects occur at multiple levels of speech analysis, including low-level acoustic level and a more complex abstract level (Samuel, 1989). The lexical nature of our test stimuli may have resulted in higher-level representations playing a greater role in the adaptation effects. Even more critically, unlike Ades (1974) and Samuel (1989) who used a single stimulus as an adaptor (e.g. *ba*), our adaptor phonemes where embedded in a variety of words which may also have contributed to adaptation effects at a more abstract level of processing. In addition, whether or not our procedure is better suited for accessing abstract phoneme representations, the important point to emphasize is that the previous authors relied on null results in their adaptation studies to reject phonemes. By contrast, we obtained robust cross-position adaptation effects in 6 out of 6 conditions, including between stop adaptors in final-syllable position spoken by a female and stop adaptors in initial position spoken by a male. Our findings are also consistent with a number of number of speech perception studies that have provided evidence for phonemes (e.g., Cutting & Day, 1975; Kazanina, Phillips, & Idsardi,2006; Morais et al., 1987; Toscano et al., 2013).

How important is it to correctly characterize the sub-lexical coding scheme for theories spoken word identification? If it turns out that phonemes contribute to speech perception, does it have broad theoretical implications? Almost certainly it does, as nicely illustrated in the domain of visual word identification. Here, the question of how written letters within words are coded for order has been the focus of a highly active research programme (for review see Grainger, 2008). Most classic models of visual word identification and naming included some version of “slot-coding” in which letters are bound to position in long-term memory (e.g., DOG is coded as D-in-position-1, O-in-position-2, and G-in-position-3; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; McClelland & Rumelhart, 1981). However, more recently, a number of detailed alternative input coding schemes have been advanced. On “open-bigram” coding schemes, letter order is represented by a set of unordered bigram units that code for the relative order of two (not necessarily adjacent) letters. For instance, DOG is accessed by the unordered set of bigrams DO, DG, and OG whereas GOD is accessed through GO, OD, and GD (e.g., Grainger & Whitney, 2004). On another view, words are accessed via position independent letter codes (e.g, the same ‘D’ letter unit is activated DOG and GOD). In order to distinguish DOG from GOD these letter codes are ordered dynamically in short-term memory (Davis 1999, 2010). For instance, in Davis (1999), a primacy gradient of level of activation is used to code for order, with the sequence D-O-G encoded through activating the ‘D’ unit the strongest, followed by ‘O’ and then ‘G’ whereas the sequence G-O-D encoded through activating the ‘G’ unit the most, followed by ‘O’, followed by ‘D’.

The critical point for present purposes is that the choice of a letter-coding scheme has had profound implications for theories of visual word identification and naming. For instance, open bigrams are well suited for recognising the visual forms of words but poorly suited for converting letters to sounds (an unordered set of DO, OG, and DG bigrams does not lend itself to naming DOG sub-lexically). For this reason, Grainger and Ziegler (2011) were forced to develop a model of word processing that includes two qualitatively different sets of representations and processes that support these two functions. By contrast, in the case of position-invariant letter codes, an entirely different set of issues arises. For example, new short-term memory processes that are capable of specifying order dynamically need to be introduced. This in turn impacts on the processes that mediate word identification and word learning (e.g., Davis, 1999, 2010). The different letter coding schemes also lead to very different empirical predictions. For instance, on slot-coding schemes, there is no similarly between OWL and HOWL (given that OWL is coded with the letters O1, W2, and L3 whereas HOWL is coded with the non-overlapping set of letters H1, O2, W3, and L4) whereas according to both open-bigram and position-invariant letter coding schemes, the two words are similar (e.g., the open bigrams OW, WL, and OL are included in both OWL and HOWL). Consistent with the latter theories, embedded words (e.g., the OWL in HOWL) are activated during visual word identification (e.g., Bowers, Davis, & Hanley, 2005; Nation & Cocksey, 2009).

To date there has been relatively little theoretical work concerned with how sub-lexical speech units are ordered, and even less empirical work directed at assessing the different hypotheses. But similar solutions have been considered. As noted above, one of the most common approaches is to assume that segments in long-term memory are context-specific, such as $[t^h]$ and [t˺] which are used in the syllable-initial and syllable-final positions respectively. In this case, word identification may be mediated by allophones whose acoustics often specify their syllable position (e.g., Mitterer et al, 2013; Pierrehumbert, 2003). On another approach, an unordered set of position-invariant (adjacent or non-adjacent) diphones is used to code for order (Hannagan et al., 2013). For example, the diphones DO, OG, and DG (along with the phonemes D, O, and G) are used to access the word representation DOG (as opposed to GOD). And yet other models include position-invariant phonemes that are dynamically bound to syllable or word position during speech perception and learning (e.g., Page & Norris, 2009).

The important point to emphasize is that these different approaches to coding the order of sub-lexical speech units entail very different theories of speech perception (just as the different approaches to ordering letters entails very different theories of visual word identification). We take our findings to lend some support to the claim that position-invariant phonemes play a role in speech perception, but clearly more empirical work is required before any strong conclusions are warranted. We hope our results will help kick-start more research on this fundamental issue that has largely been ignored in the speech perception literature.

References:

Ades, A. E. (1974). How phonetic is selective adaptation? Experiments on syllable position and vowel environment. Perception & Psychophysics, 16(1), 61-66.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Automatic semantic activation of embedded words: Is there a "hat" in "that"? *Journal of Memory and Language*, *52*(1), 131-143.

Bowers, J. S., & Michita, Y. (1998). An investigation into the structure and acquisition of orthographic knowledge: Evidence from cross-script Kanji-Hiragana priming. *Psychonomic Bulletin & Review, 5*(2), 259-264.

Bowers, J. S., Vigliocco, G., & Haan, R. (1998). Orthographic, phonological, and articulatory contributions to masked letter and word priming. *Journal of Experimental Psychology: Human Perception and Performance, 24*(6), 1705-1719.

Chomsky, N., & Halle, M. *The Sound Pattern of English. New York: Harper & Row.*

Coltheart, M. (1981). Disorders of reading and their implications for models of normal reading. Visible Language, 3, 245-286

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. doi:10.1037/0033-295X.108.1.204

Cutler, A. (2008). The 34th Sir Frederick Bartlett Lecture The abstract representations in speech processing. *Quarterly Journal of Experimental Psychology*, *61*(11), 1601–1619. doi:10.1080/13803390802218542

Cutting, J. E., & Day, R. S. (1975). The perception of stop-liquid clusters in phonological fusion. *Journal of Phonetics*, *3*(2), 99-113.

Dahan, D., & Mead, R. L. (2010). Context-Conditioned Generalization in Adaptation to Distorted Speech. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(3), 704–728. doi:10.1037/a0017449

Damian, M. F., & Dumay, N. (2009). Exploring phonological encoding through repeated segments. *Language and Cognitive Processes*, *24*(5), 685–712. doi:10.1080/01690960802351260

Dandurand, F., Hannagan, T., & Grainger, J. (2013). Computational models of location-invariant orthographic processing. *Connection Science*, *25*(1), 1–26. doi:10.1080/09540091.2013.801934

Davis, C.J. (1999). *The Self-organising Lexical Acquisition and Recognition (SOLAR) Model of Visual Word Recognition*. Doctoral dissertation, University of New South Wales, Dissertation Abstracts International, 62 (1-B), 594.

Davis, C. J. (2010). The Spatial Coding Model of Visual Word Identification. *Psychological Review*, *117*(3), 713–758. doi:10.1037/a0019738

Dell, G. S. (2014). Phonemes and production. *Language, Cognition and Neuroscience*, *29*(1), 30-32.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*, 224–238.

Fromkin, V.A. (1974). What Tips of Slung Can Tell Us About Production and Comprehension of Speech. *Journal of the Acoustical Society of America, 55*, S42–S42.

Fujimura, O. (1976). Syllables as concatenated demisyllables and affixes. Journal of the Acoustical Society ofAmerica, S55.

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology:*

Learning, memory, and cognition, *22*(5), 1166-1183.

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, *31*(3), 305-320.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes*, *23*(1), 1–35. doi:10.1080/01690960701578013

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in psychology*, *2*. doi: 10.3389/fpsyg.2011.00054

Grainger, J., & Whitney, C. (2004). Does the huamn mnid raed wrods as a wlohe? *Trends in cognitive sciences*, *8*(2), 58-59.

Hannagan, T., Magnuson, J. S., & Grainger, J. (2013). Spoken word recognition without a TRACE. *Front. Psychol.* 4 :563. doi:10.3389/fpsyg.2013.00563.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience, 29, 2-20.*

Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*, *18*(5), 943–950. doi:10.3758/s13423-011-0129-2

Johnson, K. (2011). *Acoustic and Auditory Phonetics, 3rd edition*, Wiley-Blackwell.

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences, 103*(30), 11381-11386

Keetels, M., Pecoraro, M., & Vroomen, J. (2015). Recalibration of auditory phonemes by lipread speech is ear-specific. *Cognition, 141*, 121-126.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51, 141-178.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1-15.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics, 62*(3), 615–625. doi:10.3758/BF03212113

Massaro, D. W. (1975). Preperceptual images, processing time, and perceptual units in speech perception. In Massaro, D. W. (Ed.). *Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics*. Academic Press, pp. 125-150.

McClelland, J. L. (1976). Preliminary letter identification in the perception of words and nonwords. Journal of Experimental Psychology: Human Perception and Performance, 3, 80-91.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. doi:10.1016/0010-0285(86)90015-0

McClelland, J.L., & Rumelhart, D.E. (1981). An Interactive Activation Model of Context Effects in Letter Perception .1. An Account of Basic Findings. *Psychological Review*, *88*(5), 375–407.

Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition*, *129*(2), 356–361. doi:10.1016/j.cognition.2013.07.011

Morais, J., Castro, S. L., Scliar-Cabral, L., Kolinsky, R., & Content, A. (1987). The effects of literacy on the recognition of dichotic words. *The Quarterly Journal of Experimental Psychology*, *39*(3), 451-465.

Nation, K., & Cocksey, J. (2009). Beginning readers activate semantics from sub-word orthography. *Cognition*, *110*(2), 273-278.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, *85*(3), 172–191. doi:10.1037/0033-295X.85.3.172

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms.*Philosophical Transactions of the Royal Society B: Biological Sciences*,*364*(1536), 3737-3753.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, *46*(2-3), 115-154.

Pufahl, A., & Samuel, A. G. (2014). How lexical is the lexicon? Evidence for integrated auditory memory representations. *Cognitive Psychology*, *70*, 1-30.

Reetz, H., and Jongman, A. (2009). Phonetics: Transcription, Production, Acoustics and Perception. Wiley-Blackwell.

Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of phonetics*, *45*, 91-105.

Rosenberg, A. E., Rabiner, L. R., Wilpon, J. G., & Kahn, D. (1983). Demisyllable-based isolated word recognition system. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *31*(3), 713-726.

Samuel, A. G. (1989). Insights From a Failure of Selective Adaptation - Syllable-Initial and Syllable-Final Consonants Are Different. *Perception & Psychophysics*, *45*(6), 485–493.

Samuel, A. G., & Newport, E. L. (1979). Adaptation of speech by nonspeech: evidence for complex acoustic cue detectors. *Journal of experimental psychology: human perception and performance*, *5*(3), 563-578.

Shattuck-Hufnagel, S. (1986). The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech.

Phonology, 3(01), 117-149.

Toscano, J. C., Anderson, N. D., & McMurray, B. (2013). Reconsidering the role of temporal order in spoken word recognition. *Psychonomic Bulletin & Review, 20*(5), 981–987. doi:10.3758/s13423-013-0417-0

Trubetzkoy, Nikolai (1969). *Principles of phonology*. Berkeley: University of California Press.

Vousden, J. I., Brown, G., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, *41*(2), 101–175.

Figure Captions:

Figure 1. a) The mean percentage of /s/ responses in Experiment 1 to the ambiguous test stimulus *?ail* after exposure to /f/ and /s/ adaptor words in the initial, medial and final conditions. **b)** The mean percentage of /d/ responses in Experiment 1 to the ambiguous test stimulus *?ump* after exposure to /b/ and /d/ adaptor words in the initial, medial and final conditions. Error bars represent a 95% confidence interval.

Figure 2. The mean percentage of /s/ and /d/ responses in Experiment 1 after excluding participants who did not find the test stimuli ambiguous. **a**) The mean percentage of /s/ responses to the ambiguous test stimulus *?ail* after exposure to /f/ and /s/ adaptor words in the initial, medial and final conditions. **b)** The mean percentage of /d/ responses to the ambiguous test stimulus *?ump* after exposure to /b/ and /d/ adaptor words in the initial, medial and final conditions. Error bars represent a 95% confidence interval.

Figure 3. The mean percentage of /d/ and /s/ responses in Experiment 2 to the ambiguous test stimulus ?ump and ?ail respectively after exposure to adaptor words with phonemes /b/, /d/, /f/, and /s/ in the final position. Adaptor words were spoken by a female, ambiguous test stimuli were spoken by a male. Error bars represent a 95% confidence interval.

Appendix 1. *The list of 25 adaptors used in Experiment 1 with the critical phoneme in the initial position in adaptor conditions /f/, /s/, /b/, and /d/.*

Adaptor condition			
/f/	/s/	/b/	/d/
fable	sack	bail	dainty
fact	sage	balance	dairy
faint	saint	bank	damp
fairy	sand	barn	dancer
farm	satin	basic	deep
felt	save	bath	delay
file	scar	bean	dentist
fire	score	beast	desert
flag	seem	bell	desk
flat	send	bench	device
flick	serve	berry	dial
floral	sick	bias	dice
fluid	sigh	birch	dinner
fool	silk	blame	dive
foot	silly	blaze	donkey
forget	sink	bliss	dose
formal	skin	blue	draft
forth	skull	boil	dream
foul	slam	bonus	drift
frank	slap	bother	driver
frog	smart	bowl	drug
front	smoke	brick	drum
fruit	snake	bulk	dusk
full	spirit	burst	dust
fume	swan	butter	dynamic

The list of 25 adaptors used in Experiment 1 with the critical phoneme in the beginning of the second syllable in adaptor conditions /f/, /s/, /b/, and /d/.

Adaptor condition			
/f/	/s/	/b/	/d/
affair	answer	abbey	addict
affix	arson	amber	cider
awful	bison	auburn	credit
before	blossom	cabbage	elder
cafe	classic	Cabot	Friday
crafty	cluster	cobra	giddy
defeat	consult	elbow	hardly
defend	crystal	gibbon	idol
defy	dusty	habit	ladder
effect	Easter	hobby	madly
griffin	glossy	label	meadow
inflate	gossip	limbo	media
inform	hasty	maybe	medic
leaflet	husky	orbit	modest
offend	loosen	public	needle
perfect	massive	rabbit	order
profile	message	ribbon	poodle
profit	mister	robin	puddle
profound	oyster	robot	shadow
refine	passive	rubbish	studio
refund	pistol	sober	undo
toffee	plastic	tablet	voodoo
tofu	tasty	Tibet	widow
traffic	vista	tuba	wisdom
welfare	whiskey	urban	Yiddish

The list of 25 adaptors used in Experiment 1 with the critical phoneme in the final position in adaptor conditions /f/, /s/, /b/, and /d/.

Adaptor condition			
/f/	/s/	/b/	/d/
beef	actress	Arab	ahead
belief	bass	cherub	avid
brief	boss	club	cord
calf	brass	crib	gold
cliff	chess	curb	humid
cuff	class	glib	liquid
golf	cross	globe	load
gruff	dress	grab	lucid
half	glass	grub	need
hoof	gloss	herb	orchard
knife	grass	lobe	plod
leaf	guess	perturb	pond
life	hiss	probe	reed
loaf	kiss	proverb	road
motif	lens	reverb	salad
proof	less	robe	spade
puff	loss	scrub	tend
rebuff	mass	shrub	timid
reef	mess	slob	tread
relief	miss	snob	vivid
roof	moss	superb	weed
tariff	nervous	throb	wicked
thief	pass	tribe	wind
wife	plus	tube	word
wolf	tennis	verb	yard

Appendix 2.*The list of 15 adaptors used in Experiment 2 with the critical phoneme in the final position of the word in adaptor conditions /f/, /s/, /b/, and /d/.*

Adaptor condition			
/f/	/s/	/b/	/d/
beef	bass	club	cloud
brief	boss	crab	cord
cliff	brass	curb	food
cuff	chess	herb	greed
golf	class	jib	good
gruff	cross	job	had
half	dress	knob	kid
hoof	gloss	rib	kind
leaf	guess	rob	mood
loaf	kiss	scab	mud
proof	less	shrub	nod
reef	loss	snub	road
roof	miss	sob	seed
thief	pass	verb	shed
wolf	plus	web	wild