Selective Adaptation in Speech:

Measuring the Effects of Visual and Lexical Contexts

Josh Dorsi, [1, 2] Lawrence D. Rosenblum, [1] Arthur G. Samuel, [3, 4, 5] & Serena Zadoorian[1]

[1] Department of Psychology, University of California, Riverside

[2] Department of Neurology, Pennsylvania State University, College of Medicine

[3] Department of Psychology, Stony Brook University

[4] Basque Center on Cognition Brain & Language

[5] Ikerbasque, the Basque Foundation for Science

Word Count: 9692

Author Note

Josh Dorsi is now at Penn State College of Medicine, Hershey, PA.

We have no known conflicts of interest to disclose.

Correspondence concerning this manuscript should be addressed to Josh Dorsi, Penn State College of Medicine, 30 Long Lane, Hershey, PA 17033. Email: jxd5826@psu.edu.
Support for this project was provided by NSF grant 1632530 to LDR as well as the Spanish Ministry of Science and Innovation, Grant # PSI2017-82563-P, awarded to AGS and was partially supported by the Basque Government through the BERC 2018-2021 program, and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation SEV-2015-0490 awarded to AGS.

Abstract

Speech selective adaptation is a phenomenon in which repeated presentation of a speech stimulus alters subsequent phonetic categorization. Prior work has reported that lexical, but not multisensory, context influences selective adaptation. This dissociation suggests that lexical and multisensory contexts influence speech perception through separate and independent processes (see Samuel & Lieblich, 2014). However, this dissociation is based on results reported by different studies using different stimuli. This leaves open the possibility that the divergent effects of multisensory and lexical contexts on selective adaptation may be the result of idiosyncratic differences in the stimuli rather than separate perceptual processes. The present investigation used a single stimulus set to compare the selective adaptation produced by lexical and multisensory contexts. In contrast to the apparent dissociation in the literature, we find that multisensory information can in fact support selective adaptation.

Significance Statement

This work challenges prior findings that indicate that low level speech processes precede the perceptual integration of auditory and visual information. In doing so, this research suggests that auditory and visual information are combined early in perception. This research could be beneficial for perceptual training to improve verbal communication.

Keywords

Selective Adaptation, Phonemic Restoration, Audio-Visual Integration

Selective Adaptation in Speech:

Measuring the Effects of Multisensory and Lexical Contexts

In most circumstances, individuals must perceive speech against a variety of environmental noises, such as sounds from office work, nearby traffic, and other talkers. Accurate speech perception in this complex and dynamic environment is often aided by contextual information that accompanies the auditory speech signal. This includes multisensory information, such as the visible articulations that accompany the auditory signal, as well as lexical information provided by the word in which each audible segment occurs.

Speech is both an event that occurs in the environment and also a message sent between interlocutors. That is, speech is processed both *perceptually*, to determine what articulatory event occurred, and *linguistically* to determine what meaning was conveyed by that event. That both multisensory and lexical (word context) information support speech perception is well illustrated by speech in noise listening tasks; listeners are more accurate at identifying audible speech segments when they can see the talker (Grant & Seitz, 2000; Sumby & Pollack, 1954) or when a talker is saying words as opposed to nonwords (Hirsh et al., 1954; Miller et al., 1951). But do these influences on speech identification necessarily imply that both lexical and multisensory information influence *linguistic* processing as well?

Over the last fifty years a pattern of findings associated with a phenomenon known as selective adaptation has suggested that multisensory and lexical speech information are in fact processed separately (See Samuel & Lieblich, 2014; see also Eimas & Corbit, 1973; Samuel, 2020). Selective adaptation is a finding that repeated exposure to a speech stimulus will change subsequent speech perception, such that fewer speech stimuli will be identified as belonging to the phonetic category of the previously presented item (Eimas & Corbit, 1973). For example, following 150 rapidly presented /pa/ tokens, fewer items from a /ba/-/pa/ continuum will be identified as /pa/ (Eimas & Corbit, 1973). Based on this result, Eimas and Corbit claimed that selective adaptation reflects the fatiguing of phonetic detectors. Other authors (e.g., Diehl, 1981) have argued that the shift is due to retuning phoneme classification criteria (see e.g., Kleinschmidt & Jaeger, 2016 and Samuel, 1986 for discussion of this controversy).

As a perceptual after-effect, selective adaptation has long been used as an indirect measure of perceptual processing, allowing researchers to draw inferences about perceptually-relevant features of stimuli through a task that minimizes the risk of decision bias. For example, Samuel (1997) measured selective adaptation resulting from a lexically supported speech illusion, the phonemic restoration effect. Phonemic restoration was first reported by Warren (1970), who removed a segment from an utterance and replaced this segment with noise (e.g., Warren replaced the central ‘s’ of “legi**s**latures” with a coughing sound). Warren found that in these conditions, listeners erroneously reported hearing the speech segment that had been removed. Samuel used phonemic restoration stimuli as the repetitively presented items in a selective adaptation paradigm. These phonemic restoration stimuli were words with either a central /d/ or /b/ segment that had been replaced by noise (e.g., ‘arma#ilo’ or ‘inhi#ition’). Accordingly, the test continuum on which Samuel measured adaptation was a /bI/-/dI/ test series. Samuel found that presenting noise-replaced /b/ words resulted in fewer items on the /bI/-/dI/ continuum being identified as /bI/ (/bI/ adaptation), and presenting noise-replaced /d/ words resulted in fewer items on the /bI/-/dI/ continuum being identified as /dI/ (/dI/ adaptation). In finding selective adaptation, Samuel concluded that the lexical context did not simply change the superficial identification of the replacing noise but truly phonemically restored the missing segments; that is, the lexical context supported the linguistic processing of the deleted segment.

Samuel (2001) further investigated the lexical sensitivity of selective adaptation. In this study, a single speech segment that was ambiguous between /s/ and /ʃ/ (henceforth /?/) was appended to /s/ and /ʃ/ biasing word segments such as “Tremendou” (as in “Tremendou**s**”) or “Demoli” (as in “Demoli**sh**”). These stimuli produced the classic Ganong effect (Ganong, 1980): the word context in which /?/ was inserted determined how /?/ was identified. More importantly, despite using the same /?/ segment in both conditions, Samuel (2001) found /s/ adaptation for “Tremendou/?/” and /ʃ/ adaptation for “Demoli/?/” stimuli. Samuel and Frost (2015) tested this Ganong adaptation effect in non-native English speakers. The effect was present for highly proficient bilingual participants but not for less proficient bilinguals, suggesting a link between lexical representations and selective adaptation. Thus, across three studies, selective adaptation appears to be sensitive to lexical information.

These findings contrast with what has been found for multisensory contexts, for which the McGurk effect has consistently failed to produce selective adaptation effects. The McGurk effect is the finding that certain audio-visually incongruent contexts change how the auditory component is heard (e.g., when audio ‘ba’ is dubbed with visual ‘ga’ it may be heard as ‘ga’ or ‘da’). With respect to selective adaptation, Roberts and Summerfield (1981) compared the selective adaptation effects produced by auditory-only /ba/ and /da/ to audio-visually incongruent (McGurk type) adaptors. The incongruent adaptors were composed of auditory /ba/ and visual /ga/ articulations, which generally result in /da/ percepts (MacDonald & McGurk, 1978; McGurk & MacDonald, 1976). Roberts and Summerfield (1981) found strong adaptation effects for the auditory-only /ba/ and /da/ segments. Critically, although the incongruent auditory /ba/ + visual /ga/ adaptors were frequently perceived as /da/, these authors found that the McGurk adaptors produced an adaptation effect in the same direction as the audio-only /ba/ adaptor. These results may be taken to indicate that participants had only adapted to the *auditory* token of the incongruent stimulus, with no apparent influence of the visual context or the illusory percept (but see Kleinschmidt & Jaeger, 2015, for an alternative explanation).

Concerned that the findings of Roberts and Summerfield (1981) might reflect weak audio-visual integration, Saldana and Rosenblum (1994) conducted a follow-up experiment using a more compelling McGurk stimulus (auditory /ba/ with visual /va/, which was perceived as /va/ 99% of the time). However, despite these improved stimuli, Saldana and Rosenblum (1994) replicated the original finding: adaptation appeared to be driven by the unperceived auditory stimulus (/ba/). These researchers concluded that poor cross-modal integration was unlikely to account for the results of Roberts and Summerfield (1981). Other studies have also found that that McGurk adaptors produce adaptation to the putatively unperceived auditory stimulus (Shigeno, 2002; van Linden, 2007; see also Luttke et al., 2016, and Samuel & Lieblich, 2014).

An Account with Separate Linguistic and Perceptual Processes

The success of lexical context in supporting selective adaptation, and the failure of multisensory context, is consistent with an account of speech processing with separate perceptual and linguistic speech processes. One such account is offered by Samuel and Lieblich (2014), who postulate that the perceptual process corresponds to the phenomenological experience of a speech stimulus, while the linguistic process analyzes that stimulus with respect to its role in the listener’s language. That is, the perceptual process is concerned with identifying the articulatory actions of the speaker and/or their acoustic consequences (i.e., perceiving that the talker produced a voiced bilabial action) while the linguistic process is concerned with the meaning our language assigns to that articulatory/acoustic stimulus (e.g., the bilabial place of articulation distinguishes words such as ‘bait’ & ‘date’; the voicing distinguishes words such as ‘bat’ and ‘pat’). Selective adaptation could reflect linguistic processing, independent of the perceptual process. Under the dual process account, lexical information can support both the perceptual and linguistic processes, but multisensory information can only influence the perceptual process. This hypothesis posits that some aspects of linguistically categorizing the input are not affected by multisensory information, even though this information can affect the conscious perception of the stimulus.

The proposal that linguistic processes are insensitive to multisensory information is a viable explanation for the discrepant effects of multisensory and lexical contexts on selective adaptation. However, some recent findings outside of the selective adaptation literature are less consistent with the account offered by Samuel and Lieblich (2014) (see Discussion). This led us to re-evaluate some of the selective adaptation results that originally motivated the account. To briefly re-state these selective adaptation findings: McGurk adaptors produce selective adaptation to the unperceived auditory stimulus (e.g., Roberts & Summerfield, 1981; Saldana & Rosenblum, 1994), while phonemic restoration (Samuel, 1997) and Ganong stimuli (Samuel, 2001; Samuel & Frost, 2015) support selective adaptation to a lexically-determined segment that is perceived but not present in the stimulus. These results suggest that selective adaptation follows perception when that perception is determined by lexical information, but not when it is determined by multisensory information.

However, it is also possible that these findings may reflect the fact that the multisensory selective adaptation studies have relied on the McGurk effect, in which *clear* auditory speech is *presented simultaneously* with clear and *incongruent* visual speech (McGurk & MacDonald, 1976). Indeed, accounts that view selective adaptation as a process of distributional learning, and thus predict selective adaptation from multisensory information, have noted this confound with McGurk adaptation studies (e.g., Kleinschmidt & Jaeger, 2011; 2015; 2016). Notably, the distributed learning account offered by Kleinschmidt and Jaeger (2011; 2015; 2016) strongly predicts that multisensory contexts *should* support selective adaptation. In contrast, the lexical context selective adaptation effects have been found with: a) the phonemic restoration effect (Samuel, 1997) in which the adapting phoneme is absent and replaced with noise; and b) the Ganong effect (Samuel, 2001; Samuel & Frost 2014) in which the adapting phoneme is acoustically ambiguous. In both of these cases, lexical context effects have been observed with stimuli that contain *unclear* (ambiguous) auditory segments devoid of *any simultaneous competing information*.

Thus, it could be that the failure of multisensory context to influence selective adaptation may be based on the *presence of concurrent conflicting* phonetic information, while the success of lexical context in influencing selective adaptation may be based on *unclear* phonetic information being embedded in a supportive context —and with *no conflicting* information present. Put simply, it could be that these stimulus distinctions account for the diverging effects, rather than any difference in the roles of multisensory and lexical context information[1].

The Current Study

[1] It should be noted that there are some studies that have used an ambiguous auditory + clear visual stimulus in experiments that approximate the classic selective adaptation methodology (see Baart & Vroomen, 2010; Bertelson et al., , 2003; Keetels et al., 2015; Vroomen & Baart, 2009; Vroomen et al., 2007; See also Samuel & Lieblich, 2014 for a discussion). While extensive adaptor exposure periods can produce effects similar to selective adaptation for these stimuli, in general these studies fail to find visually driven adaptation comparable to the lexical effects reported by Samuel (1997; 2001; Samuel & Frost, 2015). However, the format of ambiguous auditory + concurrent clear visual speech stimulus still retains *conflicting (concurrent)* audio-visual information and this contrasts with lexical demonstrations, which do not contain *conflicting (concurrent) information (i.e. lexical context precedes and follows the missing information)*. For this reason, the present investigation will test if the dissociation between lexical and multisensory context effects on selective adaptation is eliminated when the multisensory context lacks conflicting information.

The present investigation was designed to compare the effects of lexical and visual context on selective adaptation using *comparable* critical stimuli to test both contexts. To achieve this, we exploit the phonemic restoration effect, in which auditory information is removed from the stimulus and replaced by noise. As we noted above, these conditions result in the absence of conflicting information. The phonemic restoration method will be applied to both lexical and multisensory contexts.

In the following experiments we will measure selective adaptation effects induced by two different kinds of phonemic restoration stimuli: non-lexical multisensory phonemic restoration, and audio-only lexical phonemic restoration. The stimuli for both of these conditions originated as audio-visual recordings of a talker saying words with a central /d/ or /b/ segment (e.g. “arma**d**illo” and “inhi**b**ition”; see also Samuel, 1997). These central /b/ and /d/ segments were removed from the auditory channel and replaced with noise to produce phonemic restoration stimuli (e.g., Warren, 1970). The audio-only lexical phonemic restoration stimuli were made by removing the visual channel from these stimuli. The non-lexical multisensory restoration stimuli were made by retaining the visual channel, removing the initial and final portions of the words to produce audio-visual speech-noise-speech bi-syllables.

The critical question addressed by the following experiments is whether these lexical and multisensory restoration stimuli each can support selective adaptation effects. If selective adaptation is sensitive to a linguistic process that is insulated from multisensory information, then selective adaptation will only occur for lexical, but not multisensory, phonemic restoration contexts. If, on the other hand, the process that drives selective adaptation is also sensitive to multisensory information, then both multisensory and lexical phonemic restoration should produce selective adaptation effects.

These predictions were tested in three experiments. Experiment 1 served as a control, establishing that our full words, with no replacing noise, support selective adaptation effects (see also Samuel, 1997). In Experiment 2, the adapting segments of the words from Experiment 1 were removed and replaced with signal-correlated-noise to produce phonemic restoration stimuli. Experiment 2 had three conditions: lexical phonemic restoration (audio-only words + noise), multisensory phonemic restoration (audio-visual bi-syllables + noise), and a non-restoration control condition (*audio-only* bi-syllables + noise). There is a large literature, starting with the seminal study by Warren (1970), demonstrating phonemic restoration in audio-only words with replacing noise (see Samuel, 1996 for a review). More recently, there have been reports of phonemic restoration in audio-visual word and nonword stimuli (Abbott & Shahin, 2018; Shahin et al., 2012; Shahin & Miller, 2009; see also Jaha et al., 2020). All the stimuli used in Experiment 2 were derived from the stimuli in Experiment 1. The audio-only bi-syllables were extracted from the same stimuli used for the lexical and multisensory context conditions, and thus were appropriate control stimuli. In Experiment 3, we replicated the procedures of Experiment 2 using a different type of replacing noise.

Experiment 1

Experiment 1 began by testing the selective adaptation produced by the full word stimuli (those without any replacing noise). This experiment provides a measure of the adaptation effects when all of the acoustic speech information is available. In contrast, the subsequent experiments will assess adaptation to *illusory* speech percepts.

The results reported below came from the second iteration of this experiment; the first iteration of this experiment failed to produce selective adaptation, and being a control condition, this failed adaptation effect was puzzling. The results reported below come from an exact replication of this first iteration, using the same stimuli, procedure, sample size, and participant pool (the results of the first iteration are reported in Appendix A).

Method

Participants

Forty (16 male) University of California, Riverside students participated in Experiment 1 for course credit (Age: *M* = 19.24; *S* = 1.55). Sample size was chosen based on Samuel (1997). A power analysis found that this sample size provided our design with > 95% power to detect the selective adaptation effect reported by Samuel (1997; see Appendix B for details). All participants were native English speakers and reported normal hearing and normal or corrected to normal vision. This research was approved by the University of California, Riverside Institutional Review Board (IRB) and written consent was obtained from all participants.

Materials

All stimuli in this experiment were derived from audio-video recordings of natural words and syllables produced by a 22-year-old female speaker. This speaker was a monolingual English speaker native to Southern California. All productions were articulated at a natural pace.

Test Continuum. During audio-video recording, the speaker alternated between /da/ and /ba/ syllables, producing multiple exemplars of each. From these we selected a recording of each syllable that was judged to be the most intelligible and most prototypical of the respective category. These were used to generate the test continuum. The continuum was constructed by linear interpolation of the formant frequencies of the first three formants between the recorded /ba/ and /da/ syllables while retaining the original bandwidth contours (using a script available from http://www.mattwinn.com/praat.html; see Winn & Litovsky, 2015). The natural syllables served as endpoints of the continuum and had the onset values of (/da/: F1: 495hz; F2: 1820hz; F3: 3494) and (/ba/: F1: 652hz; F2: 1105hz; F3: 2622).

Adaptation Stimuli. The adaptation stimuli consisted of the audio channel of audio-visual recordings of words of three or more syllables with /d/ or /b/ segments in the middle of the utterance. These words were "Recon**d**ition," "Arma**d**illo," "Confi**d**ential," "Aca**d**emic," "Psyche**d**elic," "Canni**b**al," "Alpha**b**et," "Cere**b**ellum," "Cari**bb**ean," and "Inhi**b**ition". These were the same words used by Samuel (1997); the only exception being that we substituted "Cannibal" for "Exhibition" as we were concerned that the critical adaptation information (the 'ibi') of "Exhibition" may be too visually similar to "Inhibition", a factor that was relevant for Experiments 2 and 3, which relied on modifications of these stimuli.

Procedure

Each participant was alternately assigned to either the /b/ adaptor (20 participants) or the /d/ adaptor (20 participants) condition (see Dias et al., 2016 who also used a between participants adaptation comparison). In the first part of the experiment, participants made their initial baseline judgments of the tokens in the /ba/-/da/ test continuum. During this portion of the experiment, participants listened to the test items, one at a time, and for each item, reported either /da/ or /ba/ by pressing one of two labeled buttons on a computer keyboard. The test items were presented to the participants in a random order for 44 complete cycles of 8 continuum items (Eimas & Corbit, 1973; Samuel, 1986; Vroomen et al., 2007).

Following the baseline measurement, participants completed the adaptation part of the experiment. This part included 44 cycles which alternated between two phases. In the first phase of each cycle, participants were presented with a continuous stream of the auditory-only adaptor word stimuli (either /b/ or /d/) presented in a random order at a rate of approximately one word per 1.5 seconds (word length influenced the item to item duration). The primary instruction to participants in this phase of the experiment was to listen to the auditory stimuli. Additionally, during this phase of the experiment, a white dot was displayed on the screen during a randomly selected 25% of the adapting words. Participants were instructed to press the spacebar on a computer keyboard when they saw this dot. The dot monitoring task was included for consistency with Experiment 2, in which a similar methodology was used to encourage participants to attend to the visual component of the adaptors.

The content of the adaptation stream depended on the condition—/d/ or /b/ segment adaptation—to which the participant was assigned. Participants in the /d/ condition heard adapting words containing /d/ segments (e.g. Recon**d**ition, Arma**d**illo, etc.), while participants in the /b/ condition heard adapting words containing /b/ segments (e.g. Inhi**b**ition, Canni**b**al, etc.). In both conditions, the adaptation stream presented the adaptor words in a random order with the constraint that no word be repeated until all the other words in that condition had been presented.

Following this adaptation phase, each cycle included an identification phase in which participants identified all eight test continuum syllables presented in a random order. Participants indicated their responses by pressing buttons labeled “Ba” or “Da.” This portion of the experiment was identical to the baseline measure except that it consisted of only a single cycle of the test-continuum.

The first adaptation cycle included 60 adaptor words, whereas all following adaptation cycles consisted of 40 adaptor words (Samuel, 1997). There were 44 adaptation cycles, with the experimental session lasting about 70 minutes in total (Eimas & Corbit, 1973; Samuel, 1986; Vroomen et al., 2007).

A research assistant provided all instructions verbally, and these instructions were also presented as text on the computer screen during the experiment. Instructions were administered at the start of the experiment and again before the first adaptation phase began.

Results

We analyzed our results using a series of mixed effect logistic regressions (Breslow et al., 1993; Jaeger, 2008). We included a random intercept of subject (see Llompart & Casillas, 2016). We further included a random intercept of continuum item[2]. We used test block (baseline vs. post adaptation; coded as -1 & 1 respectively) as a fixed effect predicting the identification of each continuum item (coded as: [Ba]-0.5, -0.375, -0.25, -0.125, +0.125, +0.25, +0.375, +0.5 [Da]) as /ba/ or /da/ (/ba/ responses coded as 1; /da/ responses coded as 0). A computer error resulted in 2% of response data being lost from 1 participant in the /d/ adaptor condition. To understand the effects of each adaptor category we ran separate analyses testing the effect of test block in the /b/-adaptor and /d/-adaptor groups. To test for selective adaptation we analyzed the interaction of block (baseline vs. adaptation) and adaptor group (/b/-adaptors vs. /d/-adaptors; 1 & -1 respectively).

To visualize our results we tabulated the proportion of /ba/ identifications during the baseline and adaptation blocks. As can be seen in Figure 1, these /b/ and /d/ full word adaptors produced opposing identification shifts between the baseline and post adaptation (test) blocks.

[2] A reviewer pointed out that another analysis strategy would be to use continuum item as a fixed effect. This approach would enable inferences concerning which continuum items were more affected by our adaptor contexts. While this sort of question is interesting and worthy of investigation, it is outside the aims of the current investigation which sought to determine if there was any adaptation effect at all.

As Samuel (1997) also reports, the identification shift was larger in the /d/ adaptation condition than in the /b/ adaptation condition (/b/-adaptors: $\hat{\beta}$ = -0.02, *SE* = 0.03, *z* = -0.58, *p* =.56; /d/-adaptors: $\hat{\beta}$ = 0.23, *SE* = 0.03, *z* = 7.92, *p* <.001).

Next we tested if these differing adaptation shifts were statistically reliable. We found a significant interaction between experiment phase (baseline vs. test) and adaptor type (/b/-words vs. /d/-words), $\hat{\beta}$ = -0.12, *SE* = 0.02, *z* = -6.06, *p* <.001, indicating that the identification shift from baseline was different for the two adaptor contexts. These results replicate the results reported by Samuel (1997) and validate that our full word stimuli can support selective adaptation.

Experiment 2

Experiment 2 investigated if the adaptor stimuli of Experiment 1 would continue to support selective adaptation when the critical /b/ and /d/ segments were replaced by noise, in both audio-visual and lexical contexts. This experiment included three conditions: audio-visual bi-syllables, audio-only words, and audio-only bi-syllables. In each of these stimulus types, noise replaced the adapting audio /b/ or /d/ segments.

The audio-only words provided lexical, but not multisensory, context that was expected to support phonemically-restored adaptation effects (Samuel, 1997). The audio-visual bi-syllables provided multisensory, but not lexical, context and were also expected to produce phonemic restoration (e.g., Abbott & Shahin, 2018, recently reported visually supported phonemic restoration in syllable stimuli). The question tested in this experiment is whether these multisensory restoration effects would, like lexical restoration effects, produce selective adaptation. The audio-only bi-syllables provided neither lexical nor multisensory context and were thus not expected to support phonemically restored selective adaptation. Critically, the stimuli for all three conditions were constructed from the same audio-visual recordings, making them directly comparable.

If selective adaptation is sensitive to a linguistic process that is insensitive to multisensory information, then selective adaptation will only occur for lexical, but not multisensory, phonemic restoration contexts. If, on the other hand, the process that drives selective adaptation is also sensitive to multisensory information, then both multisensory and lexical phonemic restoration effects should produce selective adaptation effects.

Method

Participants

One hundred and nineteen (79 male) University of California, Riverside students participated in Experiment 2 for course credit (Age: *M* = 19.48; *S* = 1.55). Thirty-nine participants were alternately assigned to the words with replacing noise condition (19 in the /b/ replaced condition), forty to the audio-visual bi-syllable condition (20 in the /b/ replaced condition), and forty in the audio-only bi-syllable condition (20 in the /b/ replaced condition). Sample size was determined based on the power analysis reported for Experiment 1, which assumed that an audio-visual restoration adaptation effect would be similarly sized to the one reported by Samuel (1997). All participants were native English speakers and reported normal hearing and normal or corrected to normal vision. This research was approved by the University of California, Riverside Institutional Review Board (IRB) and written consent was obtained from all participants.

Materials

The materials for this experiment consisted of the /ba/-/da/ continuum used in Experiment 1 and the audio-only /b/ and /d/ words with replacing noise, as well as audio-visual and audio-only bi-syllables with replacing noise that are described below (see also Figure 2).

Adaptation Stimuli. The adaptation stimuli were created in two phases: 1) replacing the critical adapting /b/ and /d/ segments with noise; and then 2) removing the unwanted contextual information to form the three stimulus conditions. Recall that Experiment 1 presented audio-only words that were extracted from audio-visual recordings. Using the original audio-visual recordings, we removed the /b/ and /d/ segments from the auditory channel. The duration and location of the removed segment was selected iteratively: The critical /b/ or /d/ segment was first identified by visual inspection of the waveform. This selection was checked by listening to the selected segment in isolation from the rest of the word context and confirming that it could be easily identified as /b/ or /d/. After selecting a consonant segment that could be clearly identified as /b/ or /d/, the first author listened to the portion of the word preceding the selected consonant segment. If this preceding word context sounded at all like it ended with a /b/ or /d/ the selected consonant segment was adjusted to include more of the preceding word context. This process was repeated with the word context following the selected consonant segment. If the post segment word context sounded like it contained /b/ or /d/ at its onset then the consonant segment was adjusted to include more of the following word context. This process yielded isolated segments that could be clearly identified as /b/ or /d/, and word contexts preceding and following these removed segments that had no identifiable remaining /b/ or /d/ coarticulation. A naïve research assistant listened to these stimuli and confirmed these judgments. Next, for each word, we generated a white noise segment that retained the intensity profile of the deleted /b/ or /d/

segment (i.e., signal-correlated-noise; Samuel, 1997; see Figure 2). These signal-correlated-noise segments were then inserted into the audio files for each corresponding word at the point where the removed /b/ or /d/ segment had originally been. Thus, these correlated noise segments replaced the /b/ and /d/ segments.

Following the insertion of the noise segments, we edited these audio-visual words to create lexical and multisensory phonemic restoration context stimuli (and non-restoration control stimuli). The lexical phonemic restoration stimuli were created by removing the visual channel from the words, resulting in audio-only words with noise replacing the /b/ or /d/ segments. These stimuli retained the lexical information specifying the identity of the segment replaced by noise and are comparable to those used by Samuel (1997). Accordingly, these stimuli should support lexically driven phonemically-restored adaptation effects.

The multisensory restoration stimuli were created by removing the initial and final portions of each word, so that only the replacing noise and the adjacent vowels remained (i.e., for each word the bi-syllable is indicated by the bolded segments shown here: “Rec**on#i**tion,” “Arm**a#i**llo,” “Conf**i#e**ntial,” “Ac**a#e**mic,” “Psych**e#e**lic,” “Cann**i#a**l,” “Alph**a#e**t,” “Cer**e#e**llum,” “Car**i#e**an,” and “Inh**i#i**tion”; see also Figure 2). This editing produced audio-visual bi-syllables with audio noise replacing the missing /d/ or /b/. The video of the bi-syllable articulation was retained, and two brief still images corresponding to the start and the end of the auditory bi-syllable respectively were added. The silent still images were presented for durations that made the bi-syllable stimuli correspond to the duration of the original full word utterances from which they were derived. The resulting stimulus for each adaptor thus consisted of: 1) a silent still image of the speaker’s articulatory position leading into 2) the synchronized audio and dynamic visual components of the critical bi-syllable (with signal-correlated-noise replacing the critical /b/ or /d/ segment in the audio), and 3) a silent still image of the speaker's ending articulation of the bi-syllable. The durations of components 1 and 3 were set so that these bi-syllables were the same duration as the full words. Each visual stimulus showed the talker's full face, from the crown of the head to the tops of her shoulders. Importantly, these audio-visual stimuli lacked the lexical context present in the audio-only words with noise, but instead had visual information specifying the identity of the noise-replaced segment. By omitting any conflicting crossmodal information as in the McGurk effect, these stimuli provide a more analogous test of contextual information on the phonemic restoration effect.

Finally, the non-restoration control stimuli used these same bi-syllables but removed the visual channel. Being audio-only noises based on the bi-syllables, these stimuli lacked both lexical and multisensory information and were not expected to support phonemic restoration-based adaptation effects. In this way, these audio-only bi-syllables served as control stimuli, indicating whether adapting information was present in the acoustic stimuli as opposed to the lexical or multisensory context.

Procedure

With the exception of the adapting stimuli, the procedure of this experiment was identical to what was described for Experiment 1.

Each participant was assigned to either the /b/ or /d/ adaptor condition. In the first part of the experiment, participants made their initial baseline judgments of the tokens in the /ba/-/da/ test continuum. During this portion of the experiment, participants listened to the test items, one at a time, and for each item, reported either /da/ or /ba/ by pressing one of two labeled buttons on a computer keyboard. The test items were presented to the participants in a random order for 44 complete cycles of 8 continuum items.

Following the baseline measurement, the experiment cycled between participants listening/watching a continuous stream of the adaptor stimuli for their specific condition (each presented in a random order at a rate of approximately one item per 1.5 seconds) and their identification of test syllables. During the adaptation portion of the experiment, a white dot was displayed on the screen during a randomly selected 25% of the adapting items. Participants were instructed to press the spacebar on the computer keyboard when they saw this dot. The purpose of this dot monitoring task was to encourage participants in the audio-visual bi-syllable condition to attend to the visual component of the adaptors (see Samuel & Lieblich, 2014; see also Bertelson et al., 2003).

Results

We followed the same analysis approach as was used in Experiment 1; we used the same random effects structure and effect coding detailed for Experiment 1. Likewise, for visualization, we began our analysis by tabulating the proportion of /ba/ identification on the test continuum at baseline and following adaptation. A computer error resulted in 4% of response data being lost from 1 participant in the audio-visual /b/ adaptor condition. The condition means are presented in Figures 3-5. Separate analyses were run to test the effects of lexical context (words with noise), multisensory context (audio-visual bi-syllables with noise), and no context (audio-only bi-syllables with noise) conditions. The results of these analyses are reported below.

Lexical Context Effects: Audio-Only Words with Replacing Noise

The /b/ and /d/ replaced contexts produced different adaptation effects. This pattern was the result of a non-significant shift in /ba/ identifications for the /b/-replaced stimuli ($\hat{\beta}$ = 0.02, *SE* = 0.03, *z* = 0.72, *p* =.47) and a more reliable increase in /ba/ identifications for the /d/-replaced stimuli ($\hat{\beta}$ = 0.28, *SE* = 0.03, *z* = 9.55, *p* <.001; see Figure 3). The interaction testing the identification shift difference between /b/ and /d/ contexts was statistically significant ($\hat{\beta}$ = -0.12, *SE* = 0.02, *z* = -5.90, *p* <.001), demonstrating that these conditions did in fact produce selective adaptation. This result replicates the primary finding reported by Samuel (1997); in fact, the data patterns are strikingly similar. The results confirm that lexically based phonemic restoration can support selective adaptation.

Multisensory Context Effects: Audio-Visual Bi-Syllables with Replacing Noise

As was done with the audio-only words with replacing noise, we compared the identification shifts across the /b/ ($\hat{\beta}$ = 0.05, *SE* = 0.03, *z* = 1.67, *p* =.09) and /d/ ($\hat{\beta}$ = 0.24, *SE* = 0.03, *z* = 8.71, *p* <.001) conditions to determine if the audio-visual bi-syllables with noise produced multisensory phonemic restoration selective adaptation effects. The analysis of the test phase by adaptor group interaction showed a significant effect ($\hat{\beta}$ = -0.11, *SE* = 0.02, *z* = -5.58, *p* <.001) demonstrating that our audio-visual contexts were producing the expected phonetically differing adaptation effects (see Figure 4). This is the central finding of the current study: Multisensory information *can* produce selective adaptation effects. The implications of this finding will be discussed below.

No Context: Audio-Only Bi-Syllables with Replacing Noise

Both the /b/ replaced and /d/ replaced audio-only bi-syllables produced shifts in the direction of /d/ adaptation (/b/-replaced: $\hat{\beta}$ = 0.17, *SE* = 0.03, *z* = 5.38, *p* <.001; /d/-replaced: $\hat{\beta}$ =

0.28, *SE* = 0.03, *z* = 9.24, *p* <.001; see Figure 5). Samuel (1997) found a similar uniform shift pattern for noise-replaced segments in nonword stimuli, and argued that it was epiphenomenal and attributable to the lack of any clear adaptive information. Unlike the previous study, in the current study we find an interaction indicating that these shifts were significantly different from one another ($\hat{\beta}$ = -0.07, *SE* = 0.02, *z* = -3.20, *p* =.001). That is, it seems that the noise-replaced /b/ and /d/ information was, to some extent, influencing selective adaptation (see Figure 5) even in the absence of lexical and visual contextual information. This suggests that there may have been phonetic information retained in the acoustics of our noise-replaced bi-syllable stimuli. This possibility will be addressed below.

Cross Condition Interactions:

The goal of this investigation was to determine if multisensory context could support selective adaptation. Given the surprising results for the audio-only bi-syllables, we conducted an analysis testing for an interaction between experiment phase (baseline vs. adaptation), adaptor type (/b/-replaced vs. /d/-replaced adaptors) and context type (words, audio-visual bi-syllables, audio-only bi-syllables). This analysis indicated that selective adaptation was significantly larger for lexical context (words with replacing noise) than no context (audio-only bi-syllables with replacing noise), $\hat{\beta}$ = -0.07, *SE* = 0.03, *z* = -2.32, *p* =.02, indicating that lexical context had an effect beyond what was produced by the replacing noise. This analysis also found a substantial, though non-significant difference between the adaptation effects produced by audio-only and audio-visual bi-syllables, $\hat{\beta}$ = -0.05, *SE* = 0.03, *z* = -1.81, *p* =.07. Given this marginal result we reserve judgment at this point about the difference between the control stimuli and the multi-sensory stimuli.

Discussion

The main goals of Experiment 2 were to replicate the original finding of lexically mediated selective adaptation (Samuel, 1997) and to test for an effect of multisensory mediated selective adaptation. The results of the lexically mediated adaptation test were quite similar to those reported by Samuel (1997). That study reports a difference between conditions of 8.1%, just as we find an 8.1% difference (see Figure 3). In addition, in both studies the phonetic difference was driven by the larger effect of /d/ replaced stimuli, with a 6.0% shift in Samuel (1997) and a 7.3% shift in our own study[3]. Our results provide a clear replication of the lexical selective adaptation effect.

Importantly, based on the comparison of /b/ and /d/ replaced *audio-visual bi-syllable* conditions, we have extended this original finding to multisensory contexts. The magnitude of this visual context effect appears to be comparable to the effect produced by lexical context (lexical context: $\hat{\beta}$ = -0.12 vs. visual context: $\hat{\beta}$ = -0.11). The similarity of the effects produced by audio-visual context to those produced by lexical context argues against the distinction suggested by Samuel and Lieblich (2014). This point will be elaborated upon in the General Discussion.

However, one finding in Experiment 2 calls for caution at this point: The audio-only bi-syllables with replacing noise produced the same opposing /b/ vs. /d/ identification shifts as those observed in the lexical and audio-visual context conditions. While the post-hoc cross condition analysis suggests that contextual information went beyond this acoustic adaptation, more work is

[3] Note these means are calculated from the middle four continuum items, the metric reported by Samuel (1997).

needed to determine how robust these results are in the absence of this acoustic information. This issue is addressed in Experiment 3.

Experiment 3

It is known that selective adaptation can be driven by acoustic-phonetic features. For example, amplitude-shaped white noise can produce selective adaptation on a fricative-affricate continuum (e.g., Samuel & Newport, 1979). In addition, signal-correlated-noise is known to bolster phonemic restoration effects relative to other replacing sounds, presumably because of its similarity to the replaced speech segment (Samuel, 1981). This is likely related to the fact that signal-correlated-noise can also carry some basic acoustic-phonetic information as shown by better than chance performance in phoneme identification tasks (Shannon et al., 1995).

For this reason, in Experiment 3 we replicated the conditions of Experiment 2, but instead used *fixed amplitude* white noise as the replacing sound. Fixed amplitude noise uses the same carrier signal as signal-correlated-noise. The key difference is that unlike signal-correlated-noise, the temporal intensity profile of fixed amplitude noise does not correspond to the replaced speech signal (see Figure 2). While fixed amplitude noise lacks much of the structure of signal correlated noise, it has been shown to support phonemic restoration (Samuel, 1981). Experiment 3 tests whether the phonemic restoration effects produced by fixed amplitude noise are sufficient to produce selective adaptation in the conditions tested in Experiment 2.

Method

Participants

One hundred fourteen (46 male) University of California, Riverside students participated in Experiment 3 for course credit (Age: *M* = 19.04; *S* = 1.56). Thirty-seven participants were assigned to the words with replacing noise condition (20 in the /b/ replaced), thirty-seven to the audio-visual bi-syllable condition (17 in the /b/ replaced condition), and forty in the audio-only bi-syllable condition (20 in the /b/ replaced condition). Sample size was chosen based on the same power analysis reported for Experiment 2. All participants were native English speakers and reported normal hearing and normal or corrected to normal vision. This research was approved by the University of California, Riverside Institutional Review Board (IRB) and written consent was obtained from all participants.

Materials

The materials for this experiment consisted of the /ba/-/da/ continuum, the audio-only /b/ and /d/ words with replacing noise, and audio-visual and audio-only bi-syllables with replacing noise that are described above. However, the replacing noise used in this experiment was fixed amplitude white noise of the same duration as the segment it replaced and scaled to the average intensity of the words (without noise) in which it was inserted.

Procedure

The procedures of this experiment were identical to those used in Experiment 2. Briefly, participants first provided /ba/ vs. /da/ categorizations for 44 repetitions of the 8 continuum items, before going through 44 cycles of adaptation (exposure to adapting stimuli followed by continuum member categorizations). Participants were assigned to lexical (audio-only words with noise) restoration, multisensory (audio-visual bi-syllables with noise) restoration, or non-restoration (audio-only bi-syllables with noise) adaptation conditions.

Results

As was done for Experiment 2, the data from this experiment followed the same analytic approach and factor coding as was used in Experiment 1. Likewise, we used the same random effects structure detailed for Experiment 1. As was done for Experiment 2, separate analyses were run to test the effects of lexical, multisensory, and no context conditions.

No Context: Audio-Only Bi-Syllables with Replacing Fixed-Amplitude-Noise

The central question of Experiment 3 was whether lexical and multisensory context effects on selective adaptation could occur without signal-correlated-noise. In particular, would these effects still be observed in the absence of any effect in the control stimuli? For these control items, both the /b/ and /d/ replaced audio-only bi-syllables produced shifts towards fewer /ba/ identifications at test (/b/-replaced: $\hat{\beta}$ = 0.19, *SE* = 0.03, *z* = 5.68, *p* <.001; /d/-replaced: $\hat{\beta}$ = 0.21, *SE* = 0.03, *z* = 6.94, *p* <.001; see Figure 6). Critically, there was no hint of opposing adaptation effects between the /b/ and /d/ replaced conditions ($\hat{\beta}$ = -0.01, *SE* = 0.02, *z* = -0.59, *p* =.56).

Multisensory Context: Audio-Visual Bi-Syllables with Replacing Fixed-Amplitude-Noise

For the audio-visual bi-syllable condition there were shifts for both adaptors (/b/-replaced: $\hat{\beta}$ = 0.14, *SE* = 0.03, *z* = 4.60, *p* <.001; /d/-replaced: $\hat{\beta}$ = 0.34, *SE* = 0.03, *z* = 10.88, *p* <.001; see Figure 7). Critically, there was a reliable difference between the /b/ and /d/ replaced conditions ($\hat{\beta}$ = -0.11, *SE* = 0.02, *z* = -5.31, *p* <.001). This demonstrates that the multisensory context continued to support selective adaptation, even in the absence of the supportive acoustic information from signal-correlated-noise. The absence of a significant effect for this comparison using the audio-only bi-syllables makes it unlikely that the effect with the audio-visual bi-syllables is being driven by acoustic information, thus implicating a role of the multisensory contextual information. This interpretation is further tested below.

Lexical Context: Audio-Only Words with Replacing Fixed-Amplitude-Noise

As with the audio-only bi-syllables, both the /b/ and /d/ replaced audio-only words with replacing noise produced identification shifts (/b/-replaced: $\hat{\beta}$ = 0.02, *SE* = 0.03, *z* = 0.63, *p* = .53; /d/-replaced: $\hat{\beta}$ = 0.16, *SE* = 0.03, *z* = 4.72, *p* <.001; see Figure 8). Importantly, there was an interaction between test phase identification shifts and the /b/ and /d/ replaced conditions ($\hat{\beta}$ = -0.06, *SE* = 0.02, *z* = -2.85, *p* =.004), demonstrating that these stimuli had supported selective adaptation. This is the first test of phonemic restoration selective adaptation using fixed amplitude noise. These results suggest that the effect of lexical context on selective adaptation generalizes to contexts with less informative replacing noise.

Cross Condition Interactions:

As in Experiment 2, we ran an analysis comparing the adaptation effects across the different context conditions. Here we found a reliable difference between adaptation produced by audio-visual and audio-only bi-syllables ($\hat{\beta}$ = -0.10, *SE* = 0.03, *z* = -3.37, *p* < .001), but not between words and audio-only bi-syllables ($\hat{\beta}$ = -0.05, *SE* = 0.03, *z* = -1.57, *p* = .118).

Discussion

There were several motives for conducting Experiment 3. One purpose was to replicate the critical finding of Experiment 2: Would multisensory context support phonemic restoration selective adaptation when using fixed amplitude replacing noise? The significant adaptation effect found in the multisensory context condition indicates that Experiment 3 was successful in this regard.

A second goal of Experiment 3 was to look for contextually-driven adaptation with stimuli in which the residual acoustic information was not sufficient to produce adaptation. An important finding of Experiment 3 is that, unlike Experiment 2, the audio-only bi-syllables did not produce the /b/ vs /d/ differences that are characteristic of selective adaptation. Given these results, it seems unlikely that the successful effects found for the audio-visual bi-syllables are related to information retained in the audio signal.

A final goal of Experiment 3 was to determine if the lexically mediated phonemic restoration effect on selective adaptation, first reported by Samuel (1997) and replicated here in Experiment 2, would be found when fixed-amplitude noise was used rather than signal-correlated-noise. The present experiment found significantly different adaptation effects of /b/ versus /d/, adding to the findings of Samuel (1997) and those in Experiment 2. The lexical effect was not significantly different than the effect for the control stimuli, though it should be noted that that the control stimuli did not themselves promote differential adaptation.

General Discussion

Over the last forty years, a series of selective adaptation studies have shown that lexically-driven, but not multisensory, percepts can drive selective adaptation. Based on that selective adaptation literature, and a pair of new experiments, Samuel and Lieblich (2014) argued that, relative to lexical information, multisensory information has a more limited effect on speech processing, playing a role in perception but not in linguistic encoding. Here we measured adaptation effects produced by lexical or by multisensory information to address the critical theoretical question of whether both sources of information are used both perceptually and linguistically.

Across Experiments 2 and 3 we provided two tests of selective adaptation effects of lexical and multisensory contexts that were matched with respect to the acoustic support for /b/ and /d/. Experiment 2 tested selective adaptation from lexical and multisensory /b/ and /d/ restoration using the previously-used signal correlated replacing noise, while Experiment 3 tested these conditions with fixed amplitude replacing noise. Based on our results we conclude that multisensory context *can* produce selective adaptation effects.

A recurring result in our experiments was that the /d/ conditions always produced larger shifts than the /b/ conditions. In several instances, the /b/ context conditions produced effects that were actually in the direction of /d/ adaptation; however, in the restoration instances (i.e., word and audio-visual conditions) the identification shift was always smaller than what was found for the /d/ contexts. Indeed, this pattern of weaker /b/ adaptation relative to /d/ adaptation was apparent even in the non-phonemic restoration (clear words) conditions of Experiment 1. Moreover, the effects seen for /b/-adaptors appear stable across the duration of the experiment; a post-hoc correlation between adaptation block number and mean continuum item identification for that block was clearly non-significant ($r[42] = .04$, $p = 0.4$) using the data from Experiment 1. Importantly, there are no obvious lexical features that differ for our /b/ versus our /d/ words.

All of this suggests that there may have been some aspect of the stimuli that resulted in unreliable adaptation effects specific to the /b/ adaptors. Critically, the /d/ adaptors not only produced a consistently significant shift between experiment phases for the clear words and restoration conditions, but a shift that was generally significantly larger than what was found for the corresponding /b/ adaptors. These stimuli consistently produced selective adaptation — the limitations of /b/ adaptors seem unique to those adaptors rather than general to the experiments as a whole.

With this in mind, it is worth noting the possibility that phonemically restored /b/ might simply be a weak adaptor in general (though this does not explain the results of Experiment 1). There is some converging evidence to support this speculation. First, while the vast literature on selective adaptation establishes that the size of adaptation effects varies from study to study, there are some notable instances in which the adaptation effects of /b/ were smaller than the adapting effects of /d/ (e.g., Eimas & Corbit, 1973). Second, even in our Experiment 1 which used clear (that is non-phonemically restored) /b/ and /d/ stimuli, the magnitude of the /b/ adaption effect was notably less than the /d/ adaptation effect. Third, in the only other phonemic restoration selective adaptation study in the literature (Samuel, 1997), the reported results also show a less reliable adaptation effect for /b/ replaced conditions relative to /d/ replaced conditions.

Evaluation of the Audio-Only Bi-Syllables with Replacing Noise

That the signal-correlated-noise but not the fixed-amplitude-noise bi-syllables produced adaptation effects is of some interest. One possibility is that this difference is related to an interaction between potential coarticulation in the speech segments adjacent to the replacing noise and masking of this information produced by that noise. This explanation has two requirements: First, there would have to be enough coarticulation information present in the noise-adjacent segments to support identification of the noise replaced segment (despite our efforts to remove such information), indicating that articulatory information could support adaptation. Second, this explanation requires that the fixed amplitude noise masked this information more than the signal correlated noise did, thus accounting for the differential adaptation between fixed amplitude and signal correlated noise audio-only bi-syllables. To test this possibility, we ran a small (n= 23) experiment in which participants were presented with the audio-only fixed amplitude and signal correlated noise replaced bi-syllables and were asked to report if the noise had replaced /b/ or had replaced /d/ consonants.

We found that performance on this task was above chance both overall, and when examining identification for the /b/ and /d/ replaced stimuli separately (all means were greater than 55%[4]) as revealed by two-tailed single sample t-tests (all p-values less than .005; this was true even when testing consonant-vowel segments extracted from those bi-syllables which should have less coarticulatory information). This supports the first requirement of the proposed explanation: it appears that there may have been some coarticulatory information supporting the identification of the noise replaced segments. However, we found no support for the second requirement, that noise type differentially influenced the effect of this information on phoneme recovery. That is, in no condition was there a significant difference between the signal correlated and fixed amplitude noise types (smallest p-value was .47). Thus, this proposal is unable to explain the different adaptation effects observed for signal correlated versus fixed amplitude replaced segments in the auditory-only bi-syllables.

We tentatively propose that the difference between the audio-only bi-syllable conditions may be related to the acoustic information contained in the replacing noise. That is, we speculate that the envelope shape of the signal-correlated-noise may produce some speech-like adaptation effects. This interpretation is highly speculative at this point, but could have substantial

[4] Participants were numerically, though not significantly, more accurate in identifying /b/ replaced stimuli.

implications for other studies that have used signal-correlated-noise, and thus should be tested more extensively in future work.

Implications of a Multisensory Selective Adaptation Effect

Both Experiment 2 and Experiment 3 found a significant selective adaptation effect of the audio-visual bi-syllables with replacing noise. This is the central finding of the present investigation: Multisensory context can support selective adaptation. While this finding contrasts with prior work (e.g., Roberts & Summerfield, 1981; Saladana & Rosenblum, 1994) it is worth noting that is not entirely surprising. Such an effect is predicted by early integration accounts (e.g., Rosenblum et al., 2016). Moreover, recently Kleinschmidt and Jaeger (2015; 2016) have proposed that selective adaptation is the result of distributional learning, and have argued that multisensory information should produce selective adaptation. This contention was initially supported by computational modeling work, and the results of the present research provide empirical support for predictions formed by that account.

The most direct implication for finding a selective adaptation effect of multisensory context is for the dual process account put forward by Samuel and Lieblich (2014). At the time of that publication, the dual process account provided a plausible explanation for several lines of diverging results. That is, in addition to findings from the selective adaptation literature, the explanation could account for findings reported for (1) semantic priming, (2) compensation for coarticulation, (3) and neurophysiological processing of multisensory speech. However, the hypothesis was presented as a way to try to reconcile the observed findings, rather than being an idea that was designed to be tested in the study. Since its publication, new results have emerged, several of which are relevant to this account. We review this evidence here.

New Audio-visual Evidence Relevant to the Separate Processes Account

Semantic Priming

Samuel and Lieblich (2014) note that audio-visual semantic priming results reported by Ostrand and her colleagues (2011; 2016) are generally consistent with their account. The audio-visual speech of Ostrand et al's (2016) study included McGurk words, in which the auditory stimulus and the perception of that stimulus could be two different words (e.g., audio 'bait' + visual 'date' perceived as 'date'). The essential finding of this research was that semantic priming was generated by the auditory, as opposed to the visual—and putatively *perceived*—word of the McGurk stimuli (audio 'bait' + visual 'date' primed the auditory word 'worm' but not 'calendar'; but see Dorsi et al., 2017). These results are consistent with the dual processing account: the perceptual process produced the phenomenological experience of the McGurk words—the participants perceived the visual stimulus—while the linguistic process accessed the *meaning* of the unperceived auditory component of the McGurk words.

Importantly, a recent further analysis has revealed that the perceptual identification of the McGurk prime words may not have always been based on the visual speech, as had been suggested in the Ostrand et al. (2016) report (see Dorsi, 2019). Furthermore, in work done in our lab (see Dorsi et al., 2017), we found that while semantic priming can be consistent with the auditory-word of a McGurk stimulus, it is sometimes also consistent with the visual word. Critically, whether semantic priming is consistent with the auditory or visual component of a McGurk word tends to depend on how the McGurk word is *perceived* (see Dorsi, 2019). This new evidence suggests a role for multisensory integration in linguistic processing.

Compensation for Coarticulation

It is well known that there is temporal overlap in the articulation of adjacent speech segments; talkers begin each word segment before completing the preceding segment. This coarticulation (Fowler, 2010) affects the speech signal. For example, when isolated from the word “bal**d**ing” the /d/ segment may sound more like a /g/ owing to its proximity to the preceding /l/. Compensation for coarticulation refers to a phenomenon in which the perceptual system accommodates these artifacts of coarticulation, allowing listeners to perceive the segments of the speech signal as unambiguous members of their phonetic category (e.g., Mann, 1980).

In a classic demonstration of compensation for coarticulation, more items from a /ta/-/ka/ continuum are identified as /ka/ when preceded by /s/, while more are identified as /ta/ when preceded by /ʃ/ (Mann & Repp, 1980). Similar to selective adaptation, there is evidence of lexical context driving compensation for coarticulation (Elman & McClelland, 1988; see also Magnuson et al., 2003; Samuel & Pitt, 2003) and also evidence that visual context fails to do so (Vroomen & de Gelder, 2001). However, and importantly, after Samuel and Lieblich (2014) proposed their dual processes account, a meta-analysis (Viswanathan & Stephens, 2016) has been reported that supports a multisensory role in compensation for coarticulation (see also Fowler et al., 2000; Green & Norrix, 2001).

Neurophysiological Processing of Multisensory Speech

Initially, a series of findings concerning the audio-visual modulation of the auditory evoked N1 ERP (Besle et al., 2004; van Wassenhove et al., 2005, see also Stekelenburg & Vroomen, 2007) appeared consistent with the Samuel and Lieblich (2014) hypothesis. More recently, Baart and Samuel (2015) measured ERPs in response to audio-only, visual-only, or audio-visual words and nonwords. These authors report separate main effects for lexical context (words vs. nonwords) and multisensory contexts (audio-visual, audio-only, and visual-only speech), but no interaction between multisensory and lexical contexts. This study suggests that the brain processes multisensory and lexical information in two separate neurological processes (see also Zunini et al., 2019).

However, Basirat et al., (2018) have used the word repetition effect in a recent EEG study to examine the effects of multisensory and linguistic processes. The word repetition effect is the finding that prior processing of words, but not nonwords, facilitates subsequent processing of those same words (e.g., participants will identify a word faster the second time it is presented; e.g., Forbach et al., 1974). The P200 ERP component is known to be modulated by word repetition (e.g., Almeida & Poeppel, 2013). However, Basirat et al. (2018) found that this repetition effect on the P200 interacted with multisensory context. For the initial word presentation, audio-visual words were associated with a smaller ERP than were the audio-only words, suggesting that the visual context facilitated lexical access (see Basirat et al., 2018 for discussion). Their results indicate that the multisensory information of audio-visual speech may facilitate word processing analogously to the facilitation provided by word repetition. This finding suggests that, at least in some circumstances, a single brain process may be responsible for both multisensory and linguistic information (though an alternative explanation is that two separate processes each affect the P200).

Other Selective Adaptation Results

The results of Experiments 2 and 3 reported above indicate that multisensory contexts can support selective adaptation. These results converge with the results from two other findings from our lab. First, we found that McGurk adaptors produce parallel and opposing auditory *and visual* selective adaptation effects. In other words, when selective adaptation was measured on an auditory continuum, the auditory channel of the McGurk stimulus drove the effect, but when selective adaptation was measured on a visual continuum the visual (and perceived) channel of the McGurk stimulus drove the effect (Dorsi et al., 2021; see also Dias, 2016, who also measured selective adaptation on a visual continuum). Second, we investigated whether these contrasting auditory and visual adaptation effects might compete with each other crossmodally. In a meta-analysis that includes results from an experiment conducted in our lab, as well as from the adaptation studies cited by Samuel and Lieblich (2014), we found that while no single study reports a significant dilution effect for McGurk adaptors, there is a significant dilution effect *across* studies (Dorsi et al., 2021; see also Dias, 2016). It seems that McGurk adaptors cause a small, but consistent, reduction in selective adaptation relative to audio-only adaptors (Dorsi et al., 2021). Together with the experiments reported here, these findings suggest that selective adaptation is, in fact, sensitive to multisensory information.

Conclusion

Samuel and Lieblich (2014) suggested that there are separate linguistic and perceptual processes that operate during language processing. Under this account, the linguistic process is sensitive to lexical but not multisensory information, and this division seems to occur at the very earliest stages of speech processing. Although this hypothesis was consistent with the literature available at the time, and with their observed findings, research that has been published since Samuel and Lieblich's (2014) study calls this view into question. The experiments in the current study replicated the lexically driven adaptation effects reported by Samuel and Lieblich, but have clearly demonstrated that adaptation can also be driven by multisensory information. These results are consistent with predictions formed by early integration (e.g., Rosenblum et al., 2016) and computational accounts (e.g., Kleinschmidt & Jaeger, 2016). Together with the more recent findings in the literature, our results indicate that multisensory processing plays a role in both perceptual and linguistic encoding of speech.

References

Abbott, N. T., & Shahin, A. J. (2018). Cross-modal phonetic encoding facilitates the McGurk illusion and phonemic restoration. *Journal of Neurophysiology*, *120*(6), 2988–3000. http://doi.org/10.1152/jn.00262.2018

Almeida, D., & Poeppel, D. (2013). Word-specific repetition effects revealed by MEG and the implications for lexical access. *Brain and language*, *127*(3), 497-509.

Baart, M., & Samuel, A. G. (2015). Turning a blind eye to the lexicon: ERPs show no cross-talk between lip-read and lexical context during speech sound processing. *Journal of Memory and Language*, *85*(July). http://doi.org/10.1016/j.jml.2015.06.00

Baart, M., & Vroomen, J. (2010). Phonetic recalibration does not depend on working memory. *Experimental Brain Research*, *203*(3), 575–582. http://doi.org/10.1007/s00221-010-2264-9

Basirat, A., Brunellière, A., & Hartsuiker, R. (2018). The role of audiovisual speech in the early stages of lexical processing as revealed by the ERP word repetition effect. *Language Learning*, *68*(June), 80–101. http://doi.org/10.1111/lang.12265

Bertelson, P., Vroomen, J., & De Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, *14*(6), 592–597. http://doi.org/10.1046/j.0956-7976.2003.psci_1470.x

Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience, 20*(8), 2225–2234. http://doi.org/10.1111/j.1460-9568.2004.03670.x

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, *88*(421), 9-25.

Dias, J. W. (2016). *Crossmodal Influences in Selective Speech Adaptation* (Doctoral Dissertation). University of California, Riverside, Riverside, CA.

Dias, J. W., Cook, T. C., & Rosenblum, L. D. (2016). Influences of selective adaptation on perception of audiovisual speech. *Journal of Phonetics*, *56*, 75–84. http://doi.org/10.1016/j.wocn.2016.02.004

Diehl, R. L. (1981). Feature detectors for speech: A critical reappraisal. *Psychological Bulletin*, *89*(1), 1.

Dorsi, J., Rosenblum, L. D.,& Zadoorian, (2021). *Selective adaptation is sensitive to the McGurk effect*. Poster session at the 2021 annual meeting of Cognitive Neuroscience Society, Remote.

Dorsi, J. (2019). *Understanding how lexical and multisensory contexts support speech perception* (Doctoral dissertation, UC Riverside).

Dorsi, J., Ostrand, R., & Rosenblum, L. D. (2017, November). *What you see isn't always what you get, or is it? Re-examining semantic priming from McGurk stimuli.* Presented at the 58th Annual Meeting of the Psychonomic Society, Vancouver, BC, CANADA. [Abstract]

Abstracts of the Psychonomic Society, 22, 198.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, *4*, 99–109.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, *27*(2), 143–165. http://doi.org/10.1016/0749-596X(88)90071-X

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects in a lexical decision task. *Memory & Cognition*, *2*(2), 337-339.

Fowler, C. A. (2010). Embodied, embedded language use. *Ecological Psychology*, *22*(4), 286–303. http://doi.org/10.1080/10407413.2010.517115

Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(3), 877–888. http://doi.org/10.1037//O096-1523.26.3.877

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology. Human Perception and Performance*, *6*(1), 110–125. http://doi.org/10.1037/0096-1523.6.1.110

Grant, K. W., & Seitz, P. F. P. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, *108*(3), 1197–1208. http://doi.org/10.1121/1.422512

Green, K. P., & Norrix, L. W. (2001). Perception of /r/ and /l/ in a stop cluster: Evidence of cross-modal context effects. *Journal of Experimental Psychology. Human Perception and Performance*, *27*(1), 166–177. http://doi.org/10.1037/0096-1523.27.1.166

Green, P., & Macleod, C. J. (2016). *SIMR : an R package for power analysis of generalized linear mixed models by simulation*. *i*, 493–498. https://doi.org/10.1111/2041-210X.12504

Hirsh, I. J., Reynolds, E. G., & Joseph, M. (1954). Intelligibility of different speech materials. *The Journal of the Acoustical Society of America*, *26*(4), 530–538. http://doi.org/10.1121/1.1907370

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), 434-446.

Jaha, N., Shen, S., Kerlin, J. R., & Shahin, A. J. (2020). *Visual Enhancement of Relevant Speech in a ' Cocktail Party .' 33*, 277–294. https://doi.org/10.1163/22134808-20191423

Keetels, M., Pecoraro, M., & Vroomen, J. (2015). Recalibration of auditory phonemes by lipread speech is ear-specific. *Cognition*, 1–17.

Kleinschmidt, D. F., & Jaeger, T. F. (2016). Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning? *Psychonomic Bulletin and Review*, *23*(3), 678–691. http://doi.org/10.3758/s13423-015-0943-z

Kleinschmidt, D. F., & Florian Jaeger, T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. https://doi.org/10.1037/a0038695

Kleinschmidt, D., & Jaeger, T. F. (2011, June). A Bayesian belief updating model of phonetic recalibration and selective adaptation. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics* (pp. 10-19).

Llompart, M., & Casillas, J. V. (2016). Lexically driven selective adaptation by ambiguous auditory stimuli occurs after limited exposure to adaptors. *The Journal of the Acoustical Society of America*, *139*(5), EL172-EL177.

Lüttke, C. S., Ekman, M., van Gerven, M. A. J., & de Lange, F. P. (2016). McGurk illusion recalibrates subsequent auditory perception. *Scientific Reports*, *6*(August), 32891. http://doi.org/10.1038/srep32891

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, *24*(3), 253–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/704285

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: A tale of two systems? *Cognitive Science*, *27*(5), 801–805. http://doi.org/10.1016/S0364-0213(03)00067-3

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*. http://doi.org/10.3758/BF03204884

Mann, V. A, & Repp, B. H. (1980). Influence of vocalic context on perception of the [zh]-[s] distinction. *Perception & Psychophysics*. http://doi.org/10.3758/BF03204377

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264,* 746–748.

Miller, G. A, Heise, G. A, & Lighten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *The Journal of Experimental Psychology*, *41*(5), 329–335.

Ostrand, R., Blumstein, S. E., Ferreira, V. S., & Morgan, J. L. (2016). What you see isn't always what you get: Auditory word signals trump consciously perceived words in lexical access. *Cognition*, *151*, 96–107. http://doi.org/10.1016/j.cognition.2016.02.019

Ostrand, R., Blumstein, S. E., & Morgan, J. L. (2011). When hearing lips and seeing voices becomes perceiving speech: Auditory-visual integration in lexical access. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*, 1376–1381. http://doi.org/10.1016/j.neuroimage.2010.12.063.Discrete

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics*, *30*(4), 309–314. http://doi.org/10.3758/BF03206144

Rosenblum, L. D., Dias, J. W., & Dorsi, J. (2016). The supramodal brain: Implications for auditory perception. *Journal of Cognitive Psychology*, *5911*, 1–23. http://doi.org/10.1080/20445911.2016.1181691

Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *The Journal of the Acoustical Society of America*, *95*(6), 3658–3661. http://doi.org/10.1121/1.409935

Samuel, A. G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology. Human Perception and Performance*, *7*(5), 1124–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6457110

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology*, *18*(4), 452–99. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3769426

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*, *125*(1), 28–51. https://doi.org/10.1037//0096-3445.125.1.28

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, *127*(2), 97–127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, *12*(4), 348–51.

Samuel, A. G. (2020). Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language*, *111*, 104070.

Samuel, A. G., & Frost, R. (2015). Lexical support for phonetic perception during nonnative spoken word recognition. *Psychonomic Bulletin & Review*, (1970).

http://doi.org/10.3758/s13423-015-0847-y

Samuel, A. G., & Lieblich, J. (2014). Visual speech acts differently than lexical context in supporting speech perception. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(4), 1479–90. http://doi.org/10.1037/a0036656

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. http://doi.org/10.1037/h0078136

Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language, 48,* 416–434. http://doi.org/10.1016/S0749-596X(02)00514-4

Shahin, A. J., Kerlin, J. R., Bhat, J., & Miller, L. M. (2012). Neural restoration of degraded audiovisual speech. *NeuroImage*, *60*(1), 530–538. https://doi.org/10.1016/j.neuroimage.2011.11.097

Shahin, A. J., & Miller, L. M. (2009). Multisensory integration enhances phonemic restoration. *The Journal of the Acoustical Society of America*, *125,* 1744–1750. https://doi.org/10.1121/1.3075576

Shannon, R. V, Zeng, F., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primaily temporal cues. *Science*, *270*(5234), 303–304.

Shigeno, S. (2002). Anchoring effects in audiovisual speech perception. *Journal of the*

Acoustical Society of America, *111*(6), 2853–2861. http://doi.org/10.1121/1.1474446

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, *19*(12), 1964–1973. http://doi.org/10.1162/jocn.2007.91213

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*(2), 212–215.

van Linden, S. (2007). Recalibration by auditory phoneme perception by lipread and lexical information (Doctoral thesis). Tilburg University, Tilburg, The Netherlands. ISBN:978-90–5335-122–2.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–1186. http://doi.org/10.1073/pnas.0408949102

Viswanathan, N., & Stephens, J. D. W. (2016). Compensation for visually specified coarticulation in liquid–stop contexts. *Attention, Perception, & Psychophysics*. http://doi.org/10.3758/s13414-016-1187-3

Vroomen, J., & Baart, M. (2009). Recalibration of phonetic categories by lipread speech: Measuring aftereffects after a 24-hour delay. *Language and Speech*, *52*(2–3), 341–350. http://doi.org/10.1177/0023830909103178

Vroomen, J., & de Gelder, B. (2001). Lipreading and the compensation for coarticulation mechanism Vroomen de Gelder 2001.pdf. *Language and Cognitive Processes*, *16*(5/6), 661–672.

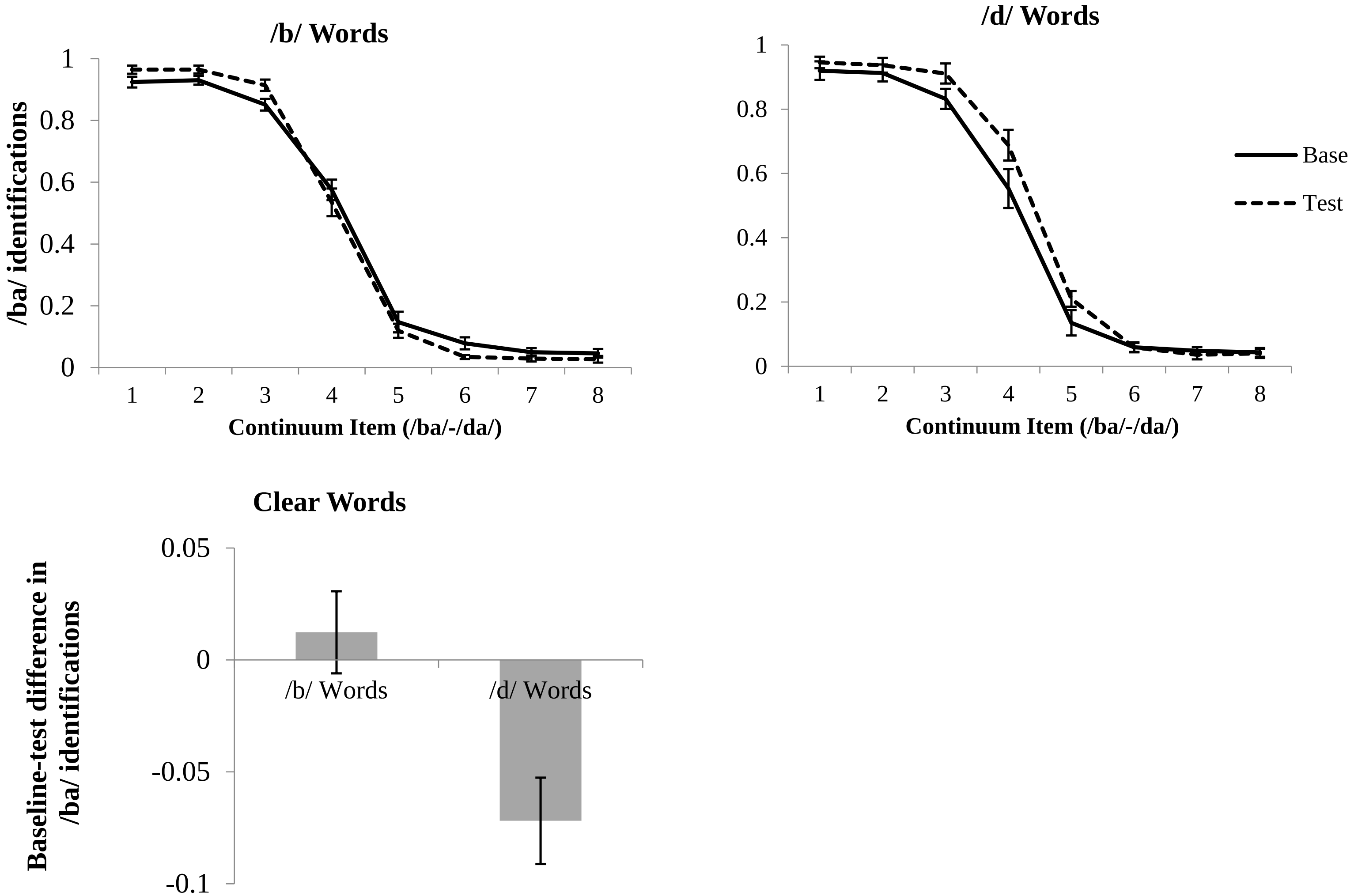
Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia*, *45*(3), 572–577. http://doi.org/10.1016/j.neuropsychologia.2006.01.031

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392–393. http://doi.org/10.1126/science.167.3917.392

Winn, M. B., & Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *Journal of the Acoustical Society of America*, *137*(3), 1430–1442. https://doi.org/10.1121/1.4908308

Zunini, R. A. L., Baart, M., Samuel, A. G., & Armstrong, B. C. (2019). Lexical access versus lexical decision processes for auditory, visual, and audiovisual items: Insights from behavioral and neural measures. *Neuropsychologia*, 107305.

Figure 1


/b/ Words
/ba/ identifications
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8
Continuum Item (/ba/-/da/)
/d/ Words
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8
Continuum Item (/ba/-/da/)
Base
Test
Clear Words
Baseline-test difference in /ba/ identifications
0.05
0
-0.05
-0.1
/b/ Words
/d/ Words

Figures 1a and 1b depict the proportion of participant “ba” identifications for each continuum item at baseline (before adaptation) and test (post adaptation). Figure 1a displays data for participants who received clear /**b**/ words during adaptation, while Figure 1b displays data for participants who received clear /**d**/ words during adaptation. Figure 1c displays the identification shifts (‘Ba’ identifications at baseline minus ‘Ba’ identifications at test) for participants of both conditions; identification shifts are averaged across the middle four continuum items for each condition. Error bars indicate standard error of the mean across subjects.

Figure 2

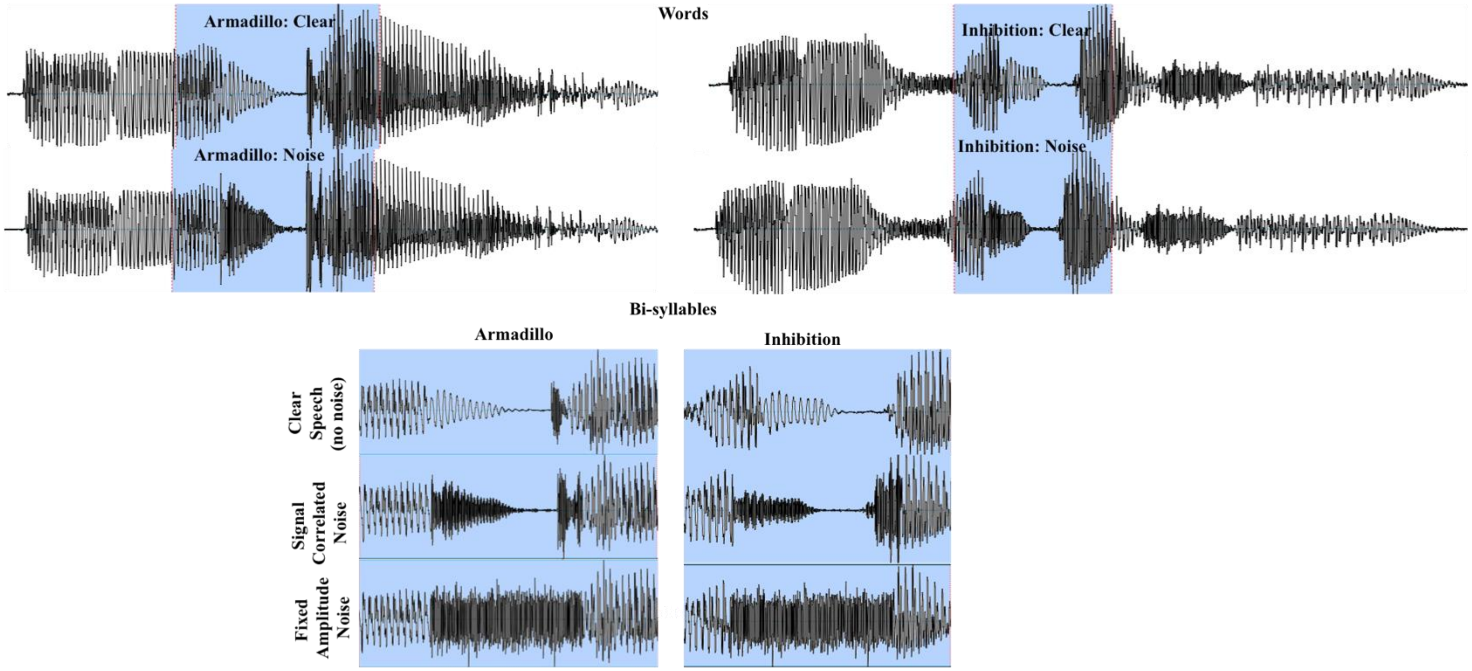
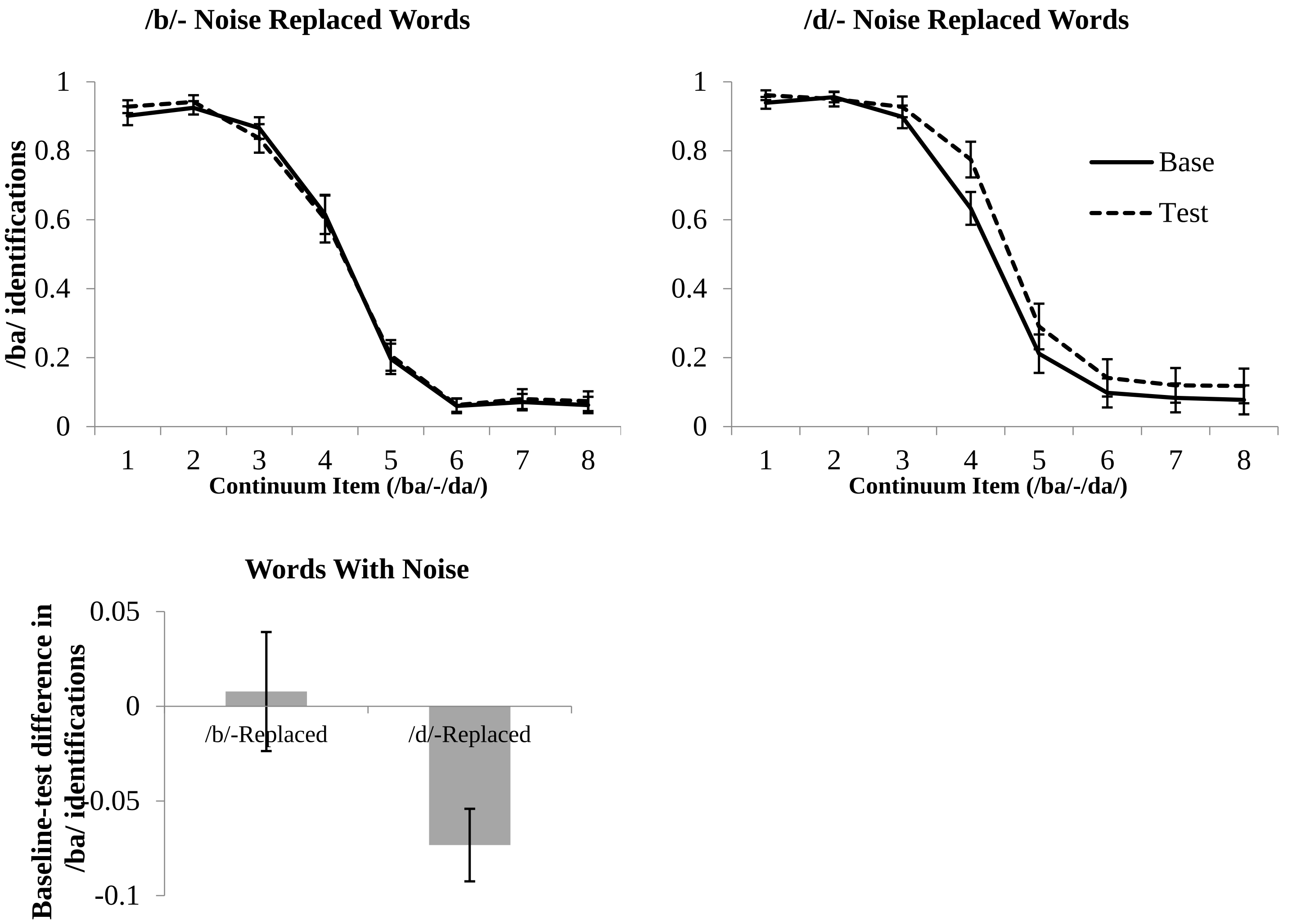

Words
Armadillo: Clear
Armadillo: Noise
Inhibition: Clear
Inhibition: Noise
Bi-syllables
Armadillo
Inhibition
Clear Speech (no noise)
Signal Correlated Noise
Fixed Amplitude Noise

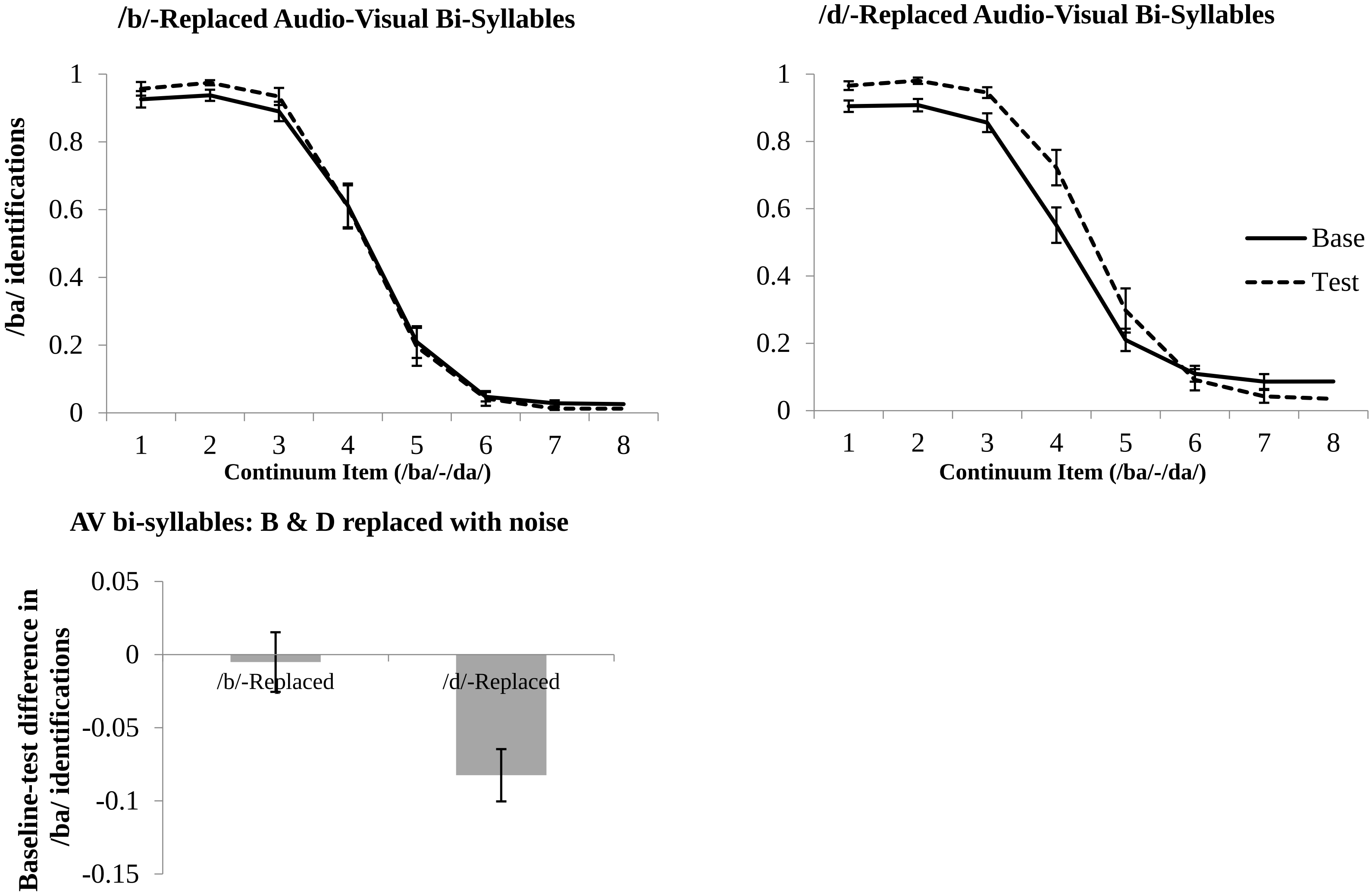
Figure 2 illustrates the auditory stimuli used in Experiments 1 - 3. The top row shows the clear, no noise, speech “Armadillo” (left) and “Inhibition” (right) used in Experiment 1. The second row shows those same words, with the adapting /d/ and /b/ segments removed and replaced with signal-correlated-noise. Note that due to coarticulation, the replacing noise includes sections of the vowels adjacent to the adapting consonant. The shaded regions denote the sections that were excised from the word context to be presented as bi-syllables. The third and fourth rows show enlargements of these sections. Note that the bi-syllables presented to participants always had replacing noise, the clear speech bi-syllables shown here are for comparison purposes only. The fifth row shows bi-syllables with non-signal correlated noise (“Fixed Amplitude Noise”) which was used in place of signal-correlated-noise during Experiment 3.

Figure 3



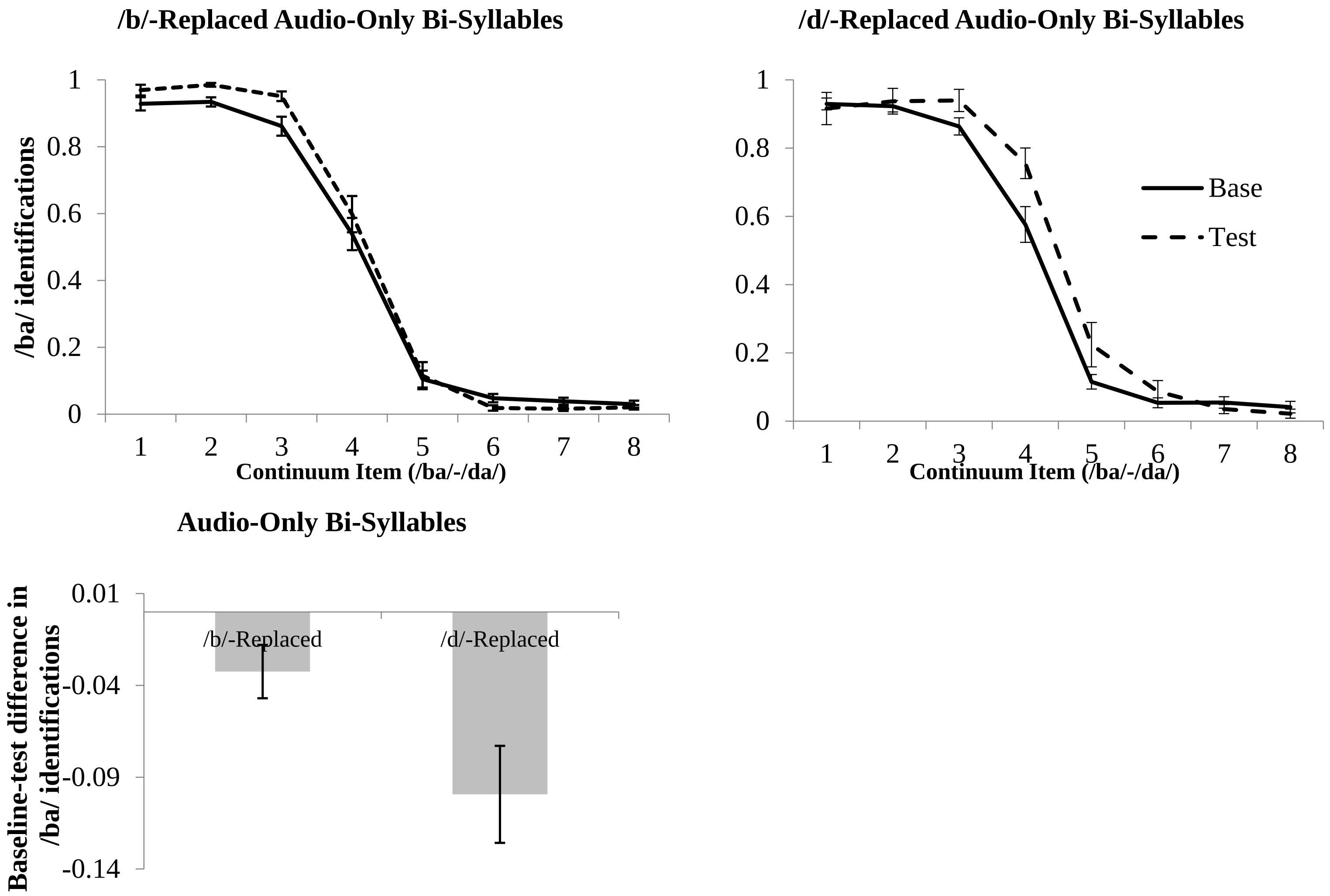
Figures 3a and 3b depict the proportion of participant "ba" identifications for each continuum item at baseline (before adaptation) and test (post adaptation). Figure 3a displays data for participants who received words with /**b**/ replaced by signal-correlated-noise during adaptation, while Figure 3b displays data for participants who received words with /**d**/ replaced by signal-correlated-noise during adaptation. Figure 3c displays the identification shifts (‘Ba’ identifications at baseline minus ‘Ba’ identifications at test) for participants of both conditions; identification shifts are averaged across the middle four continuum items for each condition. Error bars indicate standard error of the mean across subjects.

Figure 4


/b/-Replaced Audio-Visual Bi-Syllables
/ba/ identifications
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8
Continuum Item (/ba/-/da/)
/d/-Replaced Audio-Visual Bi-Syllables
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8
Continuum Item (/ba/-/da/)
Base
Test
AV bi-syllables: B & D replaced with noise
Baseline-test difference in /ba/ identifications
0.05
0
-0.05
-0.1
-0.15
/b/-Replaced
/d/-Replaced

Figures 4a and 4b depict the proportion of participant “ba” identifications for each continuum item at baseline (before adaptation) and test (post adaptation). Figure 4a displays data for participants who received audio-visual bi-syllables with /**b**/ segments replaced by signal-correlated-noise during adaptation, while Figure 4b displays data for participants who received audio-visual bi-syllables with /**d**/ segments replaced by signal-correlated-noise during adaptation. Figure 4c displays the identification shifts (‘Ba’ identifications at baseline minus ‘Ba’ identifications at test) for participants of both conditions; identification shifts are averaged across the middle four continuum items for each condition. Error bars indicate standard error of the mean across subjects.

Figure 5



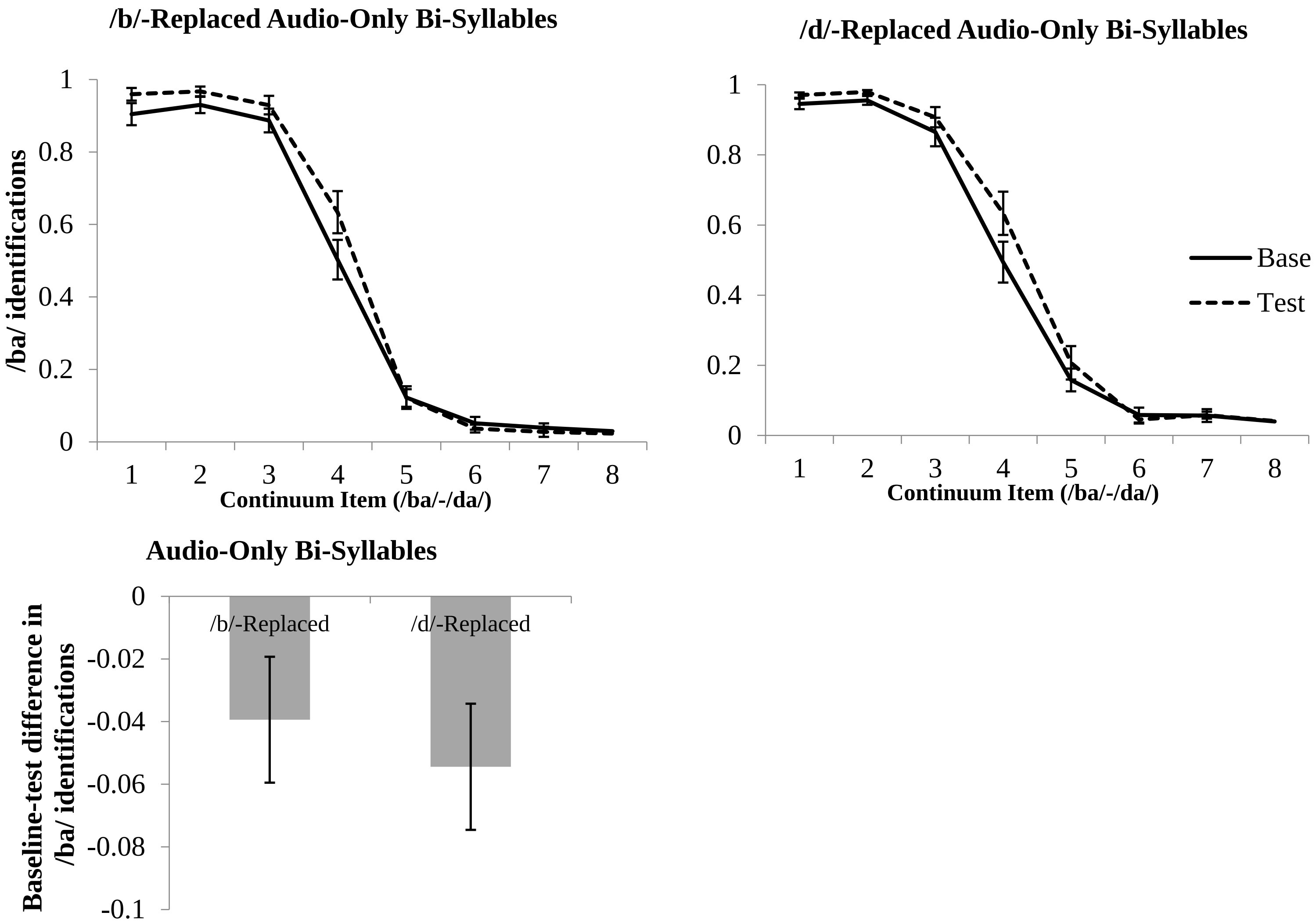
/b/-Replaced Audio-Only Bi-Syllables

/d/-Replaced Audio-Only Bi-Syllables

Audio-Only Bi-Syllables

Figures 5a and 5b depict the proportion of participant “ba” identifications for each continuum item at baseline (before adaptation) and test (post adaptation). Figure 5a displays data for participants who received audio-only bi-syllables with /**b**/ segments replaced by signal-correlated-noise during adaptation, while Figure 5b displays data for participants who received audio-only bi-syllables with /**d**/ segments replaced by signal-correlated-noise during adaptation. Figure 5c displays the identification shifts (‘Ba’ identifications at baseline minus ‘Ba’ identifications at test) for participants of both conditions; identification shifts are averaged across the middle four continuum items for each condition. Error bars indicate standard error of the mean across subjects.

Figure 6



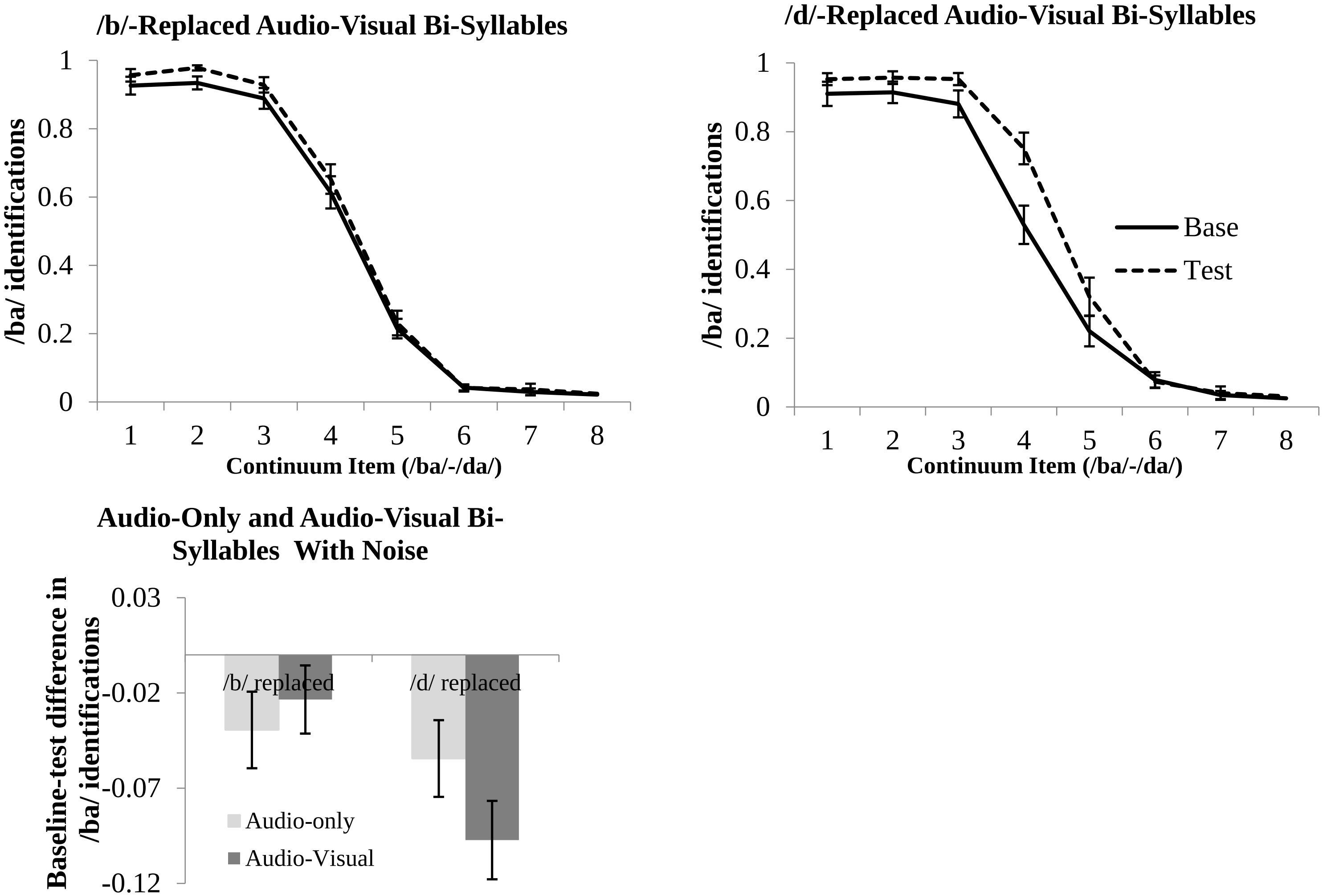
/b/-Replaced Audio-Only Bi-Syllables

/d/-Replaced Audio-Only Bi-Syllables

Audio-Only Bi-Syllables

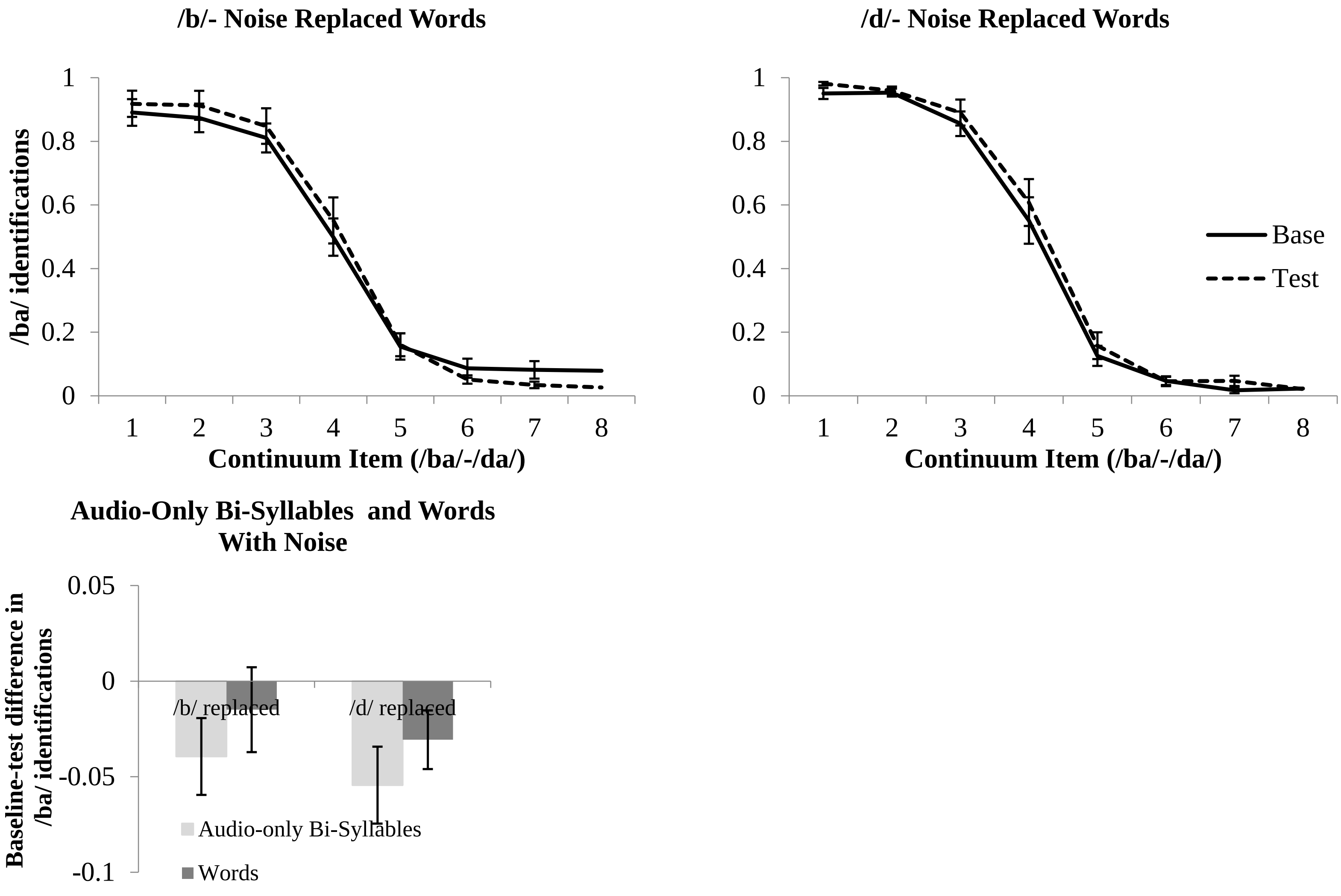
Figures 6a and 6b depict the proportion of participant “ba” identifications for each continuum item at baseline (before adaptation) and test (post adaptation). Figure 6a displays data for participants who received audio-only bi-syllables with /**b**/ segments replaced by fixed amplitude noise during adaptation, while Figure 6b displays data for participants who received audio-only bi-syllables with /**d**/ segments replaced by fixed amplitude noise during adaptation. Figure 6c displays the identification shifts (‘Ba’ identifications at baseline minus ‘Ba’ identifications at test) for participants of both conditions; identification shifts are averaged across the middle four continuum items for each condition. Error bars indicate standard error of the mean across subjects.

Figure 7



Figures 7a and 7b depict the proportion of participant “ba” identifications for each continuum item at baseline (before adaptation) and test (post adaptation). Figure 7a displays data for participants who received audio-visual bi-syllables with /**b**/ segments replaced by fixed amplitude noise during adaptation, while Figure 7b displays data for participants who received audio-visual bi-syllables with /**d**/ segments replaced by fixed amplitude noise during adaptation. Figure 7c displays the identification shifts (‘Ba’ identifications at baseline minus ‘Ba’ identifications at test) for participants of both conditions relative to corresponding audio-only bi-syllable conditions; identification shifts are averaged across the middle four continuum items for each condition. Error bars indicate standard error of the mean across subjects.

Figure 8


/b/- Noise Replaced Words
/ba/ identifications
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8
Continuum Item (/ba/-/da/)
/d/- Noise Replaced Words
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8
Continuum Item (/ba/-/da/)
Base
Test
Audio-Only Bi-Syllables and Words With Noise
Baseline-test difference in /ba/ identifications
0.05
0
-0.05
-0.1
/b/ replaced
/d/ replaced
Audio-only Bi-Syllables
Words

Figures 8a and 8b depict the proportion of participant “ba” identifications for each continuum item at baseline (before adaptation) and test (post adaptation). Figure 8a displays data for participants who received audio-only words with /**b**/ segments replaced by fixed amplitude noise during adaptation, while Figure 8b displays data for participants who received audio-only words with /**d**/ segments replaced by fixed amplitude noise during adaptation. Figure 8c displays the identification shifts (‘Ba’ identifications at baseline minus ‘Ba’ identifications at test) for participants of both conditions relative to corresponding audio-only bi-syllable conditions; identification shifts are averaged across the middle four continuum items for each condition. Error bars indicate standard error of the mean across subjects.

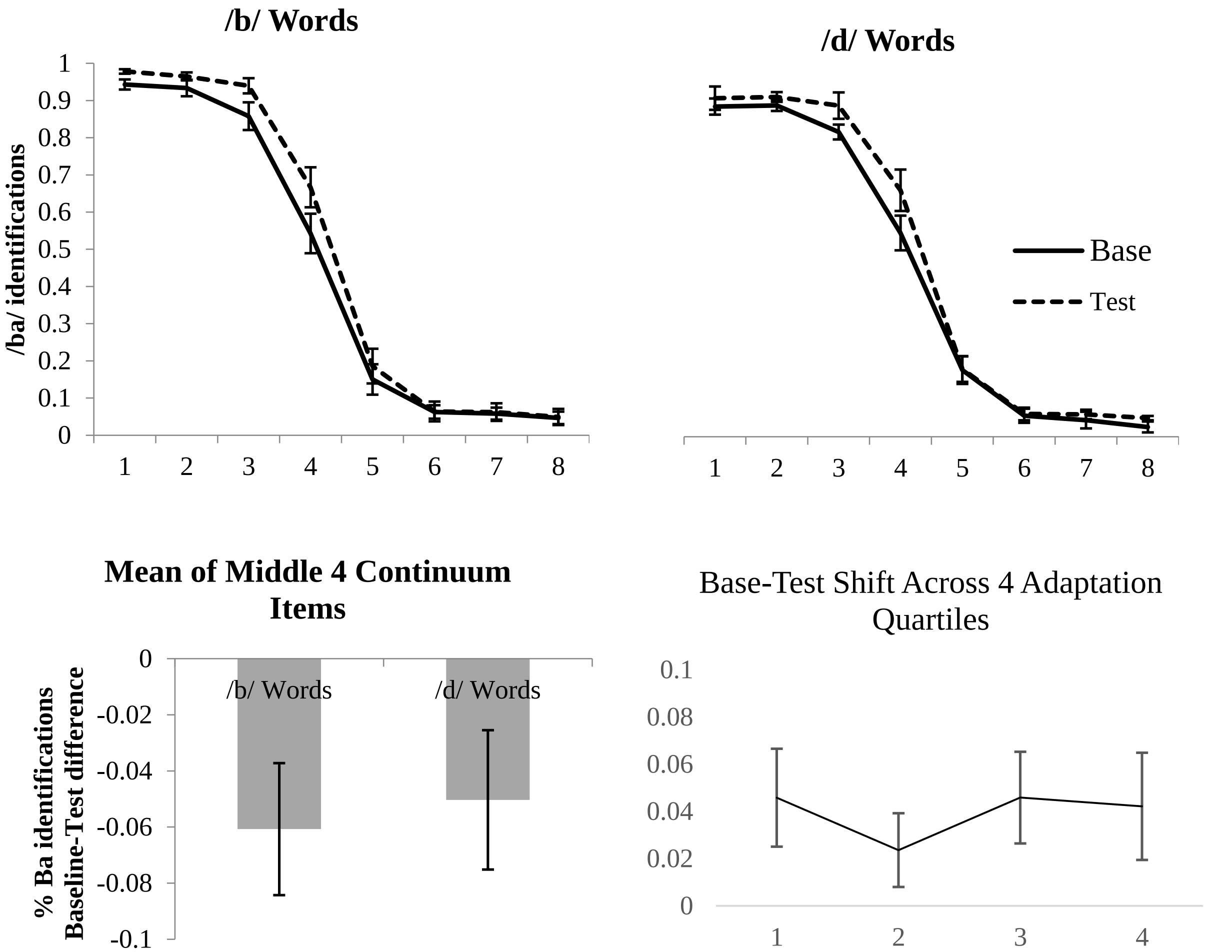
Appendix A

Results of First Run of Experiment 1

Forty (18 male) undergraduate students from University of California, Riverside participated in this experiment for course credit. Twenty of these participants were assigned to the /d/-word adaptor condition. The experiment followed the methodology detailed for Experiment 1 reported in the main text.

For the results we began by tabulating the proportion of /ba/ identifications during the baseline and adaptation blocks. As can be seen in Figures A1-A3, these /b/ and /d/ full word adaptors failed to produce the opposing baseline to adaptation identification shifts that characterize selective adaptation (/b/-adaptors: $\hat{\beta}$= 0.24, *SE* = 0.03, *z* = 8.09, *p* <.001; /d/-adaptors: $\hat{\beta}$ = 0.21, *SE* = 0.03, *z* = 7.27, *p* <.001). We tested if the /b/ and /d/ adaptation shifts were statistically dissociable. We failed to find a significant interaction between experiment phase (baseline vs. test) and adaptor type (/b/-words vs. /d/-words), $\hat{\beta}$ = 0.01, *SE* = 0.02, *z* = 0.41, *p* =.68, that would have been indicative of selective adaptation. The counter predicted identification shift for the /b/-word adaptors was remarkably stable (Figure A4) with similar sized effects occurring through the duration of the adaptation phase of the experiment; indeed, no correlation was found between the number of adaptation cycles and size of the identification effect (*r* <.01).

It is important to note that while this experiment failed to produce the interaction between experiment phase and adaptor category (/b/ vs. /d/ adaptors), the /d/ adaptors did produce a significant shift from baseline in the predicted direction. Note that the effect of /d/ adaptors for this experiment ($\hat{\beta}$ = 0.21) is quite similar to what was found for the /d/ adaptors in the main text ($\hat{\beta}$ = 0.23). In contrast, the effect of the /b/ adaptors of this experiment ($\hat{\beta}$ = 0.24) is notably different from what is reported for the same condition in the main text ($\hat{\beta}$ = -0.02). Furthermore, across all the experiments reported here, the /b/ adaptors never produced reliable adaptation effects (in contrast to the more robust effects of the /d/ adaptors); this pattern is consistent with what is reported by Samuel (1997), the study this investigation most closely matches (but see Kleinschmidt & Jaeger, [2012] and Vroomen et al. [2007] for examples of more robust /b/ adaptation). Based on this observation, it seems plausible that the absence of an adaptation effect in this experiment is the result of stochastic estimates of what is a weak/null effect /b/ adaptors.



Figures A1a and A1b depict the proportion of participant “ba” identifications for each continuum item at baseline (before adaptation) and test (post adaptation). Figure A1 displays data for participants who received clear /**b**/ words during adaptation, while Figure A2 displays data for participants who received clear /**d**/ words during adaptation. Figure A3 displays the identification shifts (‘Ba’ identifications at baseline minus ‘Ba’ identifications at test) for participants of both conditions; identification shifts are averaged across the middle four continuum items for each condition. Figure A4 shows the size of the shift from baseline during 4 quarters of the adaptation phase. Error bars indicate standard error of the mean across participants.

Appendix B

Power Analysis for Reported Experiments

The sample size for all three experiments was calculated from a power analysis that used the effect size reported for the phonemic restoration selective adaptation reported by Samuel (1997). The selective adaptation effect for full (i.e. non-phonemically restored) adaptors that is reported by Samuel (1997) was twice as large as the effect reported for the phonemic restoration adaptors. Thus, using this phonemic restoration effect size to calculate the sample size needed for our Experiment 1 (which also uses non-phonemic restoration adaptors) offers a fairly conservative estimate of our experiment's power.

Below we detail the steps to our power analysis:
Samuel (1997) reports that *the phonemically restored* /b/ adaptors produced a baseline to test (i.e. adaptation) shift of identifications of the test continuum of 2.1% while the phonemically restored /d/ adaptors produced 6.0% shift for difference of 8.1% (for a mean shift across adaptors of 1.95) between adaptor conditions (F[1, 17] =5.09); an effect equivalent to *d* = 1.09. This indicates that the test phase to adaptor category interaction used to test for selective adaptation with our mixed effect model should produce of Log Odds Ratio of 1.985 (Borenstein et al., 2009).

The standard error of the /b/ adaptor to /d/ adaptor comparison reported by Samuel (1997) is 3.59. Using *SE* = 3.59, and the mean identification shifts reported by Samuel (1997), the effect size estimate is *d*=0.547 for adaptors (/b/ vs. /d/ adaptors) and *d*=0.263 for test phase (baseline vs. test). These effects sizes were converted to Log Odds Ratio of adaptor category (0.99) and experiment phase (0.48).

To estimate power for our experiments we halved each of these estimates (to make our analysis more conservative). With these effect sizes as estimates of our fixed effects, we ran a power analysis for their interaction, assuming random intercepts for subject and continuum item (8 step ba-da). This power analysis was run using the powerCurve function from the SimR package for R (see Green & MacLeod, 2016). This function runs Monte Carlo simulations using specified parameters (i.e. the Log Odds Ratio noted above). We ran 1,000 simulations for 8 potential sample sizes (N= 1, 3, 5, 9, 15, 20, 30, 40 per group) and found that our sample size of N=20 per group provided >95% (95% CI: 99.63-100) power to detect to detect a *phonemic restoration selective adaptation effect* (i.e. Experiment 2).

As the goal of this investigation was to test whether multisensory contexts could support selective adaptation, we did not conduct an a priori power analysis for the interaction of different context adaptor types. In light of the results of Experiment 2, we felt a post-hoc analysis testing for this interaction was prudent. Using the results from that analysis we used the powerSim function of the SimR package to calculate an observed power for our test phase (baseline vs adaptation) x adaptor category (/b/-adaptors vs. /d/-adaptors) x context (audio-only bi-syllables vs. audio-visual bi-syllables) interaction, which found 58.60% power (95% CI: 54.14-62.96%) to detect the effect ($\hat{\beta}$ = -0.05). Note that the audio-only bi-syllables x audio-visual bi-syllables interaction was numerically smaller than the audio-only bi-syllables x words interaction ($\hat{\beta}$ = -0.07).

Appendix C

Experiment Results:

While the main text reports the results of mixed effects regression analyses, much of the prior research that motivated it reports the results of ANOVA and t-tests. While these analyses are inappropriate for categorical outcomes (see Jaeger, 2008) in order to facilitate comparisons to that prior literature this appendix reports the results of t-tests comparing the baseline to post adaptation identification shift, averaged across the middle four continuum items, between /b/ and /d/ type adaptor groups, the same test of selective adaptation employed by Samuel (1997). We replicate two conditions from that study, full word (Experiment 1) and words with replacing signal-correlated-noise (the lexical condition of Experiment 2). We also present the effect sizes for the data reported by Samuel (1997) for these conditions.

A reviewer pointed out that an optimal random effect structure would include random intercepts for subject and item as well as random slopes for within subject and within item manipulations (i.e. test phase, adapting context, ect.). An analysis of the lexical context effects of Experiment 2 (testing the traditional phonemic restoration effect) with this structure failed to converge. We simplified the random effects structure by removing a single random effect and re-running the analysis iteratively until a model converged. The converged model had random intercepts of subject and item, and a random slope of test phase by subject. Observed power for this analysis was only 55%. While we feel this level of power was too low to report these analyses in the main text, we do report them in this appendix for the interested reader.

Experiment 1: $t[38] = 3.17$, $p = .003$, $d = 1.03$ [2 tailed] Samuel, 1997: $d = 1.32$
$\hat{\beta} = -0.13$, $SE = 0.05$, $z = -2.72$, $p = .007$

Experiment 2: Words with SCN: $t[37] = 2.23$, $p = .032$, $d = .73$ [2 tailed] Samuel, 1997: $d = 1.09$
$\hat{\beta} = -0.12$, $SE = 0.06$, $z = -2.02$, $p = .043$

Experiment 2: AV bi-syllables with SCN: $t[38] = 2.85$, $p = .007$, $d = .93$ [2 tailed]
$\hat{\beta} = -0.11$, $SE = 0.05$, $z = -2.30$, $p = .021$

Experiment 2: AO bi-syllables with SCN: $t[38] = 2.22$, $p = .033$, $d = .72$ [2 tailed]
$\hat{\beta} = -0.08$, $SE = 0.06$, $z = -1.25$, $p = .212$

Experiment 3: AO bi-syllables with FAN: $t[35] = 0.52$, $p = .604$, $d = .18$ [2 tailed]
$\hat{\beta} = -0.01$, $SE = 0.05$, $z = -0.22$, $p = .825$

Experiment 3: AV bi-syllables with FAN: $t[35] = 2.72$, $p = .01$, $d = .92$ [2 tailed]
$\hat{\beta} = -0.11$, $SE = 0.06$, $z = -1.99$, $p = .047$

Experiment 3: Words with FAN: $t[35] = 0.56$, $p = .58$, $d = .19$ [2 tailed]
$\hat{\beta} = -0.06$, $SE = 0.06$, $z = -0.94$, $p = .35$

SCN = signal correlated noise; FAN = fixed amplitude noise.