Title

Crossmodal Influences in Selective Speech Adaptation

Permalink

https://escholarship.org/uc/item/5sd725cp

Author

Dias, James William

Publication Date

2016

Supplemental Material

https://escholarship.org/uc/item/5sd725cp#supplemental

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Crossmodal Influences in Selective Speech Adaptation

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

Doctor of Philosophy

in

Psychology

by

James William Dias

December 2016

Dissertation Committee:
Dr. Lawrence D. Rosenblum, Chairperson
Dr. Christine Chiarello
Dr. Aaron Seitz

The Dissertation of James William Dias is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

The materials in Chapter 1 constitute previously published work. The reference for this material is: 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. doi:10.1016/j.wocn.2016.02.004. These materials have been reproduced for the purposes of this dissertation with permission from the publisher.

I give esteemed thanks to my Committee Chair, Lawrence D. Rosenblum. Larry served as an incredible mentor, allowing me the freedom to explore my own research interests while providing me with the guidance to do so constructively. I learned from Larry not only the tools of the trade, but also the patience and wisdom that are invaluable characteristics of an academic advisor. I can only hope that one day I will emulate these qualities.

Esteemed thanks are also given to my dissertation committee members, Christine Chiarello and Aaron Seitz, for their advice, guidance, and support throughout my graduate program.

In large part, my success is due to the love and support of my fellow graduates. In particular, I would like to thank Kristin Layous, Z Reisz, Ho Huynh, Katie Nelson Coffey, Nicholas Shaman, Jasa Lilium, and Elysia Todd. One of the largest cohorts to grace the UCR Psychology department, ours was a group bound not only by our common goal, but also by our common love for one another. It filled me with immense joy to see

you all graduate and move on. I may be one of the last of our group to complete this leg of the journey, but I never felt left behind. Special thanks also to Elizabeth (Lizzie) McDevitt, Sarah Andrews, and Kate Sweeny. You have become a special part of my life and your love and support are invaluable to be. For all of you, I will cherish the memories of our post-stats-class Getaway retreats, our epic Halloween parties, the day-cations and vacations away from the rigors of the lab, and the countless times we simply got together to spend time with one another. I love you all.

I thank also Coleman Tuttle and Katie Heaton-Smith for your continued support through this leg in my journey. Thank you also to Graham Lelliott and the rest of my Kenpo community, who have encouraged and supported me throughout my graduate studies.

Thank you Ryan Rush, my incredibly supportive partner. Though in our pursuit of academic careers, we have not been fortunate enough to see each other on a daily basis, you are with me every day. Your love and support are felt even from a distance.

Finally, I would like to thank my mother, Ellen Deitsch. You have from my youth fostered in me a love of knowledge and a dedication to understanding. You have provided me with a wealth of love and encouragement. This dissertation, as the culmination of a doctoral degree, would not have been possible without the support you provided to me throughout my life. Thank you.

To all of you, and to all others who have contributed to my development as a researcher, as an academic, and as a human being, thank you.

ABSTRACT OF THE DISSERTATION

Crossmodal Influences in Selective Speech Adaptation

by

James William Dias

Doctorate of Philosophy, Graduate Program in Psychology
University of California, Riverside, December 2016
Dr. Lawrence D. Rosenblum, Chairperson

Repeated presentation of speech syllables can change identification of ambiguous syllables, a perceptual aftereffect known as *selective speech adaptation* (e.g., Eimas & Corbit, 1973). Adaptation to auditory speech syllables can change identification of auditory speech and adaptation to visual (lipread) syllables can change identification of visual speech. Investigations of potential crossmodal influences suggest that selective speech adaptation depends on shared sensory information between adaptors and test-stimuli (Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014). These findings have been cited as support for theories suggesting the speech process treats auditory and visual speech differently (e.g., Diehl, Kluender, & Parker, 1985; Massaro, 1987; Samuel & Lieblich, 2014). However, the lack of crossmodal influences in selective speech adaptation is inconsistent with a large literature across other speech domains suggesting crossmodal influences occur early in the speech process. These crossmodal findings from other speech domains are often cited as support for theories suggesting the speech process treats auditory and visual speech the same, based on modality-neutral information available in both (for a review, see Rosenblum, Dorsi, & Dias, 2016). This dissertation provides a rigorous investigation of crossmodal influences in auditory and visual speech adaptation. Chapter 1 investigated whether crossmodal adaptation can influence perception of visually-influenced audiovisual phonetic percepts. Chapter 2 investigated whether the subtle (non-significant) crossmodal adaptation measured previously (Roberts & Summerfield, 1981) is replicable and significant when provided sufficient statistical power (using larger sample sizes). Chapter 3 used bimodal adaptors to investigate whether crossmodal adaptation can augment changes in speech perception following within-modality adaptation (when adapted and tested in the same sensory modality). The results of the dissertation suggest that selective speech adaptation is sensitive to modality-neutral articulatory information. However, crossmodal adaptation is subtle, requiring larger samples of participants to reach statistical significance. These subtle crossmodal effects may be unobservable following concurrent within-modality adaptation, perhaps suggesting a ceiling effect when adapting and testing in the same sensory modality.

Table of Contents

List of Figures xii

List of Tables xiv

Introduction 1

Speech is multimodal 2

Selective speech adaptation 5

Lexical influences in selective speech adaptation 10

Adaptation effects outside of the speech domain 12

Purpose of the current series of investigations 14

Chapter 1 15

Chapter 2 17

Chapter 3 18

References 20

Chapter 1: Influences of Selective Adaptation on Perception of Audiovisual Speech

Introduction 27

Selective Adaptation in Speech Perception 29

The Current Investigation 34

Methods 40

Participants 40

Materials 41

Procedure 44

Results 45

Discussion 47

What Information Does Selective Adaptation Influences? 48

Conclusion 52

References 54

Chapter 2: Selective adaptation of crossmodal speech information: A case of small but consistent effects

Introduction 63

Investigations of crossmodal influences in selective speech adaptation 64

Lexical influences in selective speech adaptation 69

Summary of the problem 72

Current investigation 74

Experiment 1 75

Method 76

Participants 76

Materials 76

Procedure 78

Results & Discussion 80

Meta-Analysis 1 83

Results & Discussion 84

Experiment 2 85

Methods 85

Participants 85

Materials 86

Procedure 86

Results and Discussion 86

Meta-Analysis 2 87

Results & Discussion 88

Experiment 3 88

Methods 90

Participants 90

Materials 91

Procedure 91

Results & Discussion 91

General Discussion 93

Segment differences in crossmodal selective speech adaptation 94

Differences between within and crossmodal speech adaptation 96

Implications for theoretical accounts 99

References 103

Footnotes 110

Chapter 3: Audio-visual selective speech adaptation does not exhibit a bimodal advantage

Introduction 120

Current investigation 124

Method 126

Participants 126

Materials 126

Procedure 128

Results 128

Discussion 129

References 132

Conclusion 139

Crossmodal Influences in Selective Speech Adaptation 139

Implications for Prior Theoretical Accounts 141

References 144

List of Figures

Figure 1.1 61

The nine-item audiovisual test continuum of integrated phonetic percepts. As Gaussian Blur becomes stronger, the salience of the visual information becomes less. However, for all items in the continuum, the auditory component remains the same (/ba/). The strength of the McGurk illusion becomes weaker as visibility of the mouth decreases. As a result, greater reliance is put on the auditory component of the audiovisual stimulus, decreasing perception of the illusory “va” percept.

Figure 1.2 62

The proportion of /ba/ responses for each of the nine test-continuum items prior to (Baseline) and post adaptation (Adapted) for each of the four adaptors. The bottom left panel illustrates the phonemic boundary shift for each adaptor. Positive values denote shifts towards the /ba/-end of the continuum. Negative values denote shifts towards the /va/-end of the continuum. Error bars represented the standard error of the mean.

Figure 2.1 116

Still images take from the visual test-continuum at the place of articulation. Stimulus 1 corresponds with the unambiguous visual-/va/ video and stimulus 11 corresponds with the unambiguous visual-/ba/ video.

Figure 2.2 117

Experiment 1 test-stimulus identifications. The average proportion each continuum test-stimulus was identified as /ba/ before (Baseline, dashed line) and after (Adapted, solid line) adaptation. Data is represented for each of the eight groups, depending on tested continuum modality and adaptor. Error bars represent standard errors of the mean.

Figure 2.3 118

Experiment 2 test-stimulus identifications. The average proportion each continuum test-stimulus was identified as /ba/ before (Baseline, dashed line) and after (Adapted, solid line) adaptation. Data is represented for each of the four groups, depending on tested continuum modality and adaptor. Error bars represent standard errors of the mean.

Figure 2.4 119

Experiment 3 test-stimulus identifications. The average proportion each continuum test-stimulus was identified as /ba/ before (Baseline, dashed line) and after (Adapted, solid line) adaptation. Data is represented for each of the four groups, depending on tested continuum modality and text adaptor. Error bars represent standard errors of the mean.

Figure 3.1 138

The average proportion each continuum test-stimulus was identified as /ba/ before (Baseline, dashed line) and after (Adapted, solid line) adaptation. Data is represented for each of the eight groups, depending on tested continuum modality and adaptor. Error bars represent standard errors of the mean.

List of Tables

Table 1.1 59
Hypothesized changes in categorization of audiovisual test-stimuli following adaptation

Table 1.2 60
Phonemic boundary shifts following adaptation

Table 2.1 111
Experiment 1: Phonemic category boundaries and magnitudes of boundary shift following adaptation.

Table 2.2 112
Experiment 1: Differences between the phonemic boundary shifts following within-modality and crossmodal adaptation.

Table 2.3 113
Meta-analyses of crossmodal selective adaptation in speech.

Table 2.4 114
Experiment 2: Phonemic category boundaries and magnitudes of boundary shift following adaptation.

Table 2.5 115
Experiment 3: Phonemic category boundaries and magnitudes of boundary shift following adaptation.

Table 3.1 135
Changes in speech categorization following unimodal and bimodal speech adaptation in previous studies.

Table 3.2 136
Phonemic category boundaries and magnitudes of boundary shifts following adaptation.

Table 3.3 137
Differences between the phonemic boundary shifts following bimodal and unimodal adaptation.

Introduction

Repeated presentation of speech syllables can change identification of ambiguous syllables, a perceptual aftereffect known as *selective speech adaptation*. Adaptation to auditory speech syllables can change identification of auditory speech (e.g., Eimas & Corbit, 1973) and adaptation to visual (lipread) syllables can change identification of visual speech (e.g., Jones, Feinberg, Bestelmeyer, DeBruine, & Little, 2010). However, studies investigating the influences of visual speech adaptation on auditory speech perception have failed to find significant crossmodal adaptation (e.g., Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014). The results of these studies are inconsistent with evidence for crossmodal influences between auditory and visual speech in other domains, even at the earliest stages of perceptual processing (for a review, see Rosenblum, Dorsi, & Dias, 2016). These results are also inconsistent with evidence for crossmodal adaptation between audition and vision in other perceptual areas (e.g., motion aftereffects) (for a review, see Rosenblum, Dias, & Dorsi, 2016).

Failing to find crossmodal adaptation between visual and auditory speech has implications for understanding the underlying basis of the speech processes. Research from other speech domains supports theories of speech processing based on modality-neutral articulatory primitives (e.g., Fowler, 2004; Rosenblum, 2005; Rosenblum, Dorsi, et al., 2016). However, the lack of evidence for crossmodal selective speech adaptation may support theories that suggest auditory and visual speech information are perceptually different (e.g., Diehl, Kluender, & Parker, 1985; Diehl, Lotto, & Holt, 2004; Samuel &

Lieblich, 2014), and that speech may involve an initial level of processing that is sense-specific (e.g., Massaro, 1987, 2015).

The inconsistent evidence for crossmodal influences between selective speech adaptation and other speech/non-speech domains suggest further investigation is warranted. Further, there are characteristics of the studies that have previously investigated crossmodal influences in selective speech adaptation that warrant consideration for future investigations.

In the following sections, I will first provide a brief summary of the literature investigating multisensory influences in speech. I will next provide a brief summary of the literature on selective speech adaptation, including those studies that investigated crossmodal speech adaptation. Finally, I will summarize the purpose of the following series of investigations, including the investigation-specific hypotheses relating to crossmodal influences in selective speech adaptation.

Speech is multimodal

Research investigating the multisensory nature of speech has revealed that the speech mechanism is sensitive to information provided across sensory modalities (for a review, see Rosenblum, Dorsi, et al., 2016). Perhaps the most striking example of the multisensory nature of speech is demonstrated by the McGurk effect (McGurk & MacDonald, 1976), where a visible (lipread) syllable can change perception of a different audible syllable (for a review, see Dias, Cook, & Rosenblum, in press; Green, 1996). The visual influence is so strong that perceivers will typically identify the syllable they hear as different from the syllable identified when the auditory stream is heard in isolation

(MacDonald & McGurk, 1978; McGurk & MacDonald, 1976). In fact, visible speech information can influence auditory speech identification even when the visual information is reduced to kinematic primitives using point-light displays of articulating faces (e.g., Rosenblum & Saldaña, 1996). Similar cross-sensory influences in auditory speech perception have been demonstrated using tactile (e.g., Fowler & Dekle, 1991; Gick & Derrick, 2009) and kinesthetic (Ito, Tiede, & Ostry, 2009; Sams, Mottonen, & Sihvonen, 2005) sources of information.

Auditory speech perception can also improve when *redundant* information is provided across sensory modalities. For example, the identification of words spoken in noisy environments (e.g., white noise) is more accurate when the speaker's articulating face is visible (e.g., Erber, 1975; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954). Similarly, visibility of a speaker's articulating face can improve comprehension of accented speech (e.g., Navarra & Soto-Faraco, 2007; Sueyoshi & Hardison, 2005) and speech that conveys complicated content (Arnold & Hill, 2001; Reisberg, McLean, & Goldfield, 1987). Some have suggested that visual speech can improve comprehension of auditory speech because of the lawful relationship between heard and seen speech gestures: The articulations of the vocal tract that shape speech sounds also produce correlated changes in the visible face. As such, both the auditory and visual streams provide information pertaining to a common articulatory event. Speech may be based on these modality-neutral articulatory primitives (Fowler, 2004; Rosenblum, Dorsi, et al., 2016), as opposed to abstract representations based on auditory (e.g., Diehl & Kluender, 1989) or motor substrates (e.g., Liberman & Mattingly, 1985).

The notion that auditory and visual speech convey information for common articulatory primitives is supported by a number of other studies (for a review, see Rosenblum, Dorsi, et al., 2016). For example, perceivers are able to match a heard speaker with the correct visual display of the articulating speaker (e.g., Lachs & Pisoni, 2004a, 2004b), even when the heard and seen speech are reduced in quality of information (i.e., sine-wave acoustic transformations and visual point-light displays) (Lachs & Pisoni, 2004c). Similarly, experience lipreading a speaker can improve comprehension of that same speaker's audible speech (Rosenblum, Miller, & Sanchez, 2007), and experience hearing a speaker talk can improve lipread comprehension of the same speaker (Sanchez, Dias, & Rosenblum, 2013). These studies suggest that experience with the modality-neutral articulatory style of a speaker can be shared across sensory modalities.

The notion that speech perception is based on the articulatory information conveyed across sensory modalities is also demonstrated by the similar influences heard and seen speech can have on speech production. For example, perceivers will subtly imitate the idiosyncratic characteristics of a speaker's style of speaking after hearing (e.g., Goldinger, 1998; Pardo, 2006) and seeing (lipreading) (Miller, Sanchez, & Rosenblum, 2010) the speaker talk. In fact, perceivers will imitate a speaker more if they can hear *and* see the speaker at the same time (Dias & Rosenblum, 2011, 2015), mirroring the bimodal audio-visual advantages to speech perception discussed earlier.

Neurophysiological evidence suggests that many structures important to the speech process are sensitive to information available across the auditory and visual modalities. For example, auditory cortex activates in response to lipreading (e.g., Calvert et al., 1997; Campbell, 2008; Pekkola et al., 2005). Similarly, bimodal audiovisual stimuli with common auditory components, but different visual components, will produce different patterns of activity in auditory cortex (e.g., Callan, Callan, Kroos, & Vatikiotis-Bateson, 2001; Colin et al., 2002; Sams et al., 1991). For example, auditory cortex activity will differ in response to an audio-/pa/-visual-/pa/ stimulus and an audio-/pa/-visual-/ka/ stimulus (a combination typically perceived as /ka/ or /ta/; e.g., McGurk & MacDonald, 1976), even though the auditory components of these two audiovisual stimuli are the same (Sams et al., 1991). Some neurophysiological evidence even suggests that visual speech information can modulate activity in auditory brainstem (Musacchia, Sams, Nicol, & Kraus, 2006). The neurophysiological evidence suggests that crossmodal speech influences occur at the earliest stages of perceptual processing, perhaps even before the extraction of phonetic features (for a review, see Rosenblum, Dorsi, et al., 2016).

Despite the evidence suggesting the speech process is sensitive to information available across sensory modalities, research in the field of *selective speech adaptation* has challenged some assumptions regarding the nature of these crossmodal influences.

Selective Speech Adaptation

Selective speech adaptation describes a perceptual aftereffect where identification of ambiguous speech syllables changes following repeated exposure (adaptation) to some stimulus (e.g., Eimas & Corbit, 1973; Ganong, 1978; Goldstein & Lackner, 1972; Remez, 1980). Eimas and Corbit (1973) demonstrated the effect in a classic study by measuring changes in the identification of speech tokens along an auditory /ba/-to-/pa/ test-continuum that followed adaptation to auditory utterances of /ba/ and /pa/. More test-items were identified as /pa/ following adaptation to /ba/ and more test-items were identified as /ba/ following adaptation to /pa/. Eimas and Corbit (1973) explained the effect as resulting from a fatiguing of "linguistic feature detectors" (hypothetical cognitive constructs sensitive to speech features) that resulted from repeated exposure to phonetic information. For the example above, /ba/ and /pa/ differ only in their voice-onset time feature (/pa/ has a longer VOT). Following adaptation to /pa/, the detectors sensitive to longer voice-onset times were fatigued, causing more continuum items to be perceived as having shorter voice-onset-times, consistent with /ba/. To demonstrate this point, Eimas and Corbit (1973) measured changes in the identification of speech tokens along the same auditory /ba/-to-/pa/ test-continuum following adaptation to auditory utterances of /da/ and /ta/, which have similar voice-onset times to /ba/ and /pa/, respectively. More test-items were identified as /pa/ following adaptation to /da/ and more test-items were identified as /ba/ following adaptation to /ta/. However, subsequent investigations of selective speech adaptation suggest that auditory adaptor stimuli need not be speech in nature (e.g., white-noise) to change auditory speech identification, so long as there is spectrotemporal overlap between adaptors and test stimuli (e.g., Diehl et al., 1985; Kat & Samuel, 1984; Samuel & Newport, 1979).

The classic literature in selective speech adaptation demonstrates how auditory speech identification can change following auditory adaptation (either speech or non-speech). More recently, studies have also demonstrated that visual speech identification can change following visual speech adaptation. For example, Jones et al. (2010) found that identification of visual stimuli along an /m/-to-/u/ continuum of still-face images can change following adaptation to a still face articulating /m/ or /u/. More continuum items were identified as /m/ following adaptation to /u/ and more were identified as /u/ following adaptation to /m/. Similarly, Baart and Vroomen (2010) found that identification of stimuli along an /omso/-to-/onso/ continuum of dynamically articulating faces can change following adaptation to a video (with sound) of a speaker articulating /omso/ or /onso/.

The evidence discussed in the previous section suggesting that the speech process is sensitive to information available across sensory modalities has prompted some to investigate whether similar crossmodal influences exist in selective speech adaptation. Early on, Cooper and colleagues (Cooper, 1974; Cooper, Billings, & Cole, 1976; Cooper, Blumstein, & Nigro, 1975) observed a link between speech production and auditory speech perception in selective speech adaptation. Though the motor system is not typically considered a perceptual modality, as such, it is inherently associated with kinesthetic and tactile sensations that can guide speech production. Cooper (1974) found that /pa/ was produced with shorter voice onset times following repeated presentation of an auditory-/pa/ stimulus, making the /pa/ productions more /ba/-like in structure. In two follow-up investigations, repeated speech productions were found to change auditory speech identification (Cooper et al., 1976; Cooper et al., 1975). For example, Cooper et al. (1975) found that more test-stimuli from an auditory /ba/-to-/da/ continuum were identified as /da/ following repeated production of /ba/. These results suggest that adaptation to articulatory information available in the kinesthetic and tactile sensations associated with speech production can change perception of related articulatory information in auditory speech.

However, Roberts and Summerfield (1981) later failed to find significant changes in auditory speech identification following crossmodal adaptation of *visual* speech information. Changes in the identification of test-stimuli along an auditory /bɛ/-to-/dɛ/ continuum were observed following adaptation to auditory-/bɛ/ and auditory-/dɛ/, but not following adaptation to visual-/bɛ/ or visual-/dɛ/. They also found that changes in auditory speech identification following audiovisual adaptation (audiovisual-/bɛ/ or audiovisual-/dɛ/) were no greater than changes following auditory adaptation, suggesting that redundant information provided across modalities does not enhance selective speech adaptation.

Interestingly, Roberts and Summerfield (1981) also found that more test-stimuli were identified as /dɛ/ following adaptation to an auditory-/bɛ/-visual-/gɛ/ adaptor, an audio-visual combination that is typically identified as /dɛ/ (e.g., McGurk & MacDonald, 1976). This last result suggested that perceivers adapted to the sensory (auditory) information shared between adaptor and test-stimuli, not to the perceived phonetic percept. Roberts and Summerfield (1981) concluded from these results that selective speech adaptation depends on sense-specific (auditory) spectrotemporal overlap between adaptors and test stimuli.

Similar to Roberts and Summerfield (1981), Saldaña and Rosenblum (1994) measured changes in the identification of test-stimuli along an auditory /ba/-to-/va/

continuum following adaptation to an auditory-/ba/-visual-/va/ stimulus, an audiovisual combination that was identified as /va/ 99% of the time. Despite the compelling nature of the phonetic percept generated by the audiovisual adaptor, changes in speech identification reflected adaptation to the auditory component of the adaptor, not the phonetic percept, similar to Roberts and Summerfield (1981) (see also Samuel & Lieblich, 2014). To reconcile their results with the broader literature suggesting that the speech process is sensitive to multisensory input, Saldaña and Rosenblum (1994) suggested that selective speech adaptation may occur at a level of processing prior to the integration of auditory and visual information. In other words, selective speech adaptation may occur at a point in perceptual processing sensitive to sensory-specific input, but not yet at a point where modality-neutral articulatory primitives are extracted from sensory input.

The results of Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994) do not support theories suggesting a speech mechanism based solely on modality-neutral articulatory primitives extracted across modalities at the earliest levels of perceptual processing. Instead, these studies seem to suggest that the perceptual system first processes speech information with some discretion between the senses. These studies have since been cited as support for theories proposing a perceptual mechanism that treats auditory and visual speech information differently (e.g., Diehl et al., 1985; Samuel & Lieblich, 2014), and perhaps involves an initial level of processing that is sense-specific (e.g., Massaro, 1987).

Lexical influences in selective speech adaptation

Interestingly, a collection of studies conducted by Samuel and colleagues (Samuel, 1997, 2001; Samuel & Lieblich, 2014) suggest that changes in auditory speech identification can follow adaptation to adaptors that do *not* share sense-specific spectrotemporal information with test stimuli. In these studies, the top-down influences of lexical and linguistic knowledge can play a role in selective speech adaptation. Samuel (1981, 1987) previously observed that spoken word-utterances are typically perceived as intact after a critical consonant is removed and replaced with noise, an effect known as *phonemic restoration* (Samuel, 1981, 1987). Later, Samuel (1997, 2001) found that auditory speech identification could change following adaptation to these restored-stimuli. For example, more test-stimuli along an auditory /ba/-to-/da/ continuum were identified as /ba/ following adaptation to an auditory utterance of “armadillo” that had the /d/ removed and replaced with signal-correlated white noise (“arma-illo”). The results of these investigations suggest that spectrotemporal overlap between adaptors and test-stimuli is not required for selective speech adaptation.

Samuel and Lieblich (2014) proposed a theoretical account to reconcile why phonetic information resulting from lexical context can affect selective speech adaptation (e.g., Samuel, 1997, 2001) while phonetic information in visual speech cannot (e.g., Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994). They suggested that selective speech adaptation is primarily influenced by sense-specific bottom-up information. In the absence of bottom-up information, top-down knowledge can influence selective speech adaptation. From their theoretical perspective, Samuel and Lieblich

(2014) consider the influences of visual speech and lexical context on auditory speech perception to both be top-down influences — perhaps related to the learned associations between heard and seen speech cues or between speech sounds and the words that contain them (e.g., Diehl et al., 1985; Massaro, 1987; Samuel, 1981, 1987). Based on their findings, along with those of Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994), Samuel and Lieblich (2014) suggest that the influence of visual speech on auditory speech perception is perceptual in nature, but not linguistic. However, the influence of lexical context is *both* perceptual and linguistic. As such, lexical context can provide more information than visual speech to affect selective speech adaptation.

Though this theoretical account can rationalize the counterintuitive findings between lexical and visual speech influences in selective speech adaptation, the account does not take into consideration findings from other research areas. For example, the claim that visual speech is non-linguistic is not consistent with findings suggesting that visual speech can exhibit linguistic qualities, using both behavioral (e.g., Kim, Davis, & Krins, 2004) and neurophysiological measures (for a review, see Bernstein & Liebenthal, 2014). Also, the theoretical account offered by Samuel and Lieblich (2014) does not consider the aforementioned work by Cooper and colleagues demonstrating selective speech adaptation between auditory speech perception and (the kinesthetic and tactile sensations associated with) speech production (Cooper, 1974; Cooper et al., 1976; Cooper et al., 1975). Evidence from these and other speech domains (discussed previously) suggest auditory and visual speech perception share a common modality-neutral basis. However, the theoretical account offered by Samuel and Lieblich (2014) implies that visual and auditory speech have different bases and are associated only through conventional use (i.e., experience). Their theoretical account also suggests that auditory and visual speech are associated with different types of information: Auditory speech is associated with phonetic forms and the words that contain them, but visual speech is associated only with phonetic forms.

Despite the evidence from other speech domains for a speech mechanisms that is based on modality-neutral articulatory primitives (e.g., Fowler, 2004; Rosenblum, 2005; Rosenblum, Dorsi, et al., 2016), Samuel and Lieblich (2014) propose (based on findings in selective speech adaptation) a speech mechanism that treats bottom-up information as sense-specific and subsequently treats auditory and visual speech information differently. However, the sense-specificity observed in selective speech adaptation (Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014) is inconsistent with evidence from other perceptual areas that demonstrate crossmodal adaptation.

Adaptation effects outside of the speech domain

Other areas of research report adaptation effects similar to selective speech adaptation. Perhaps the most familiar of these adaptation effects come from the vision literature. Perceptual aftereffects are found following adaptation to visual motion (Anstis, Verstraten, & Mather, 1998; Carleson, 1962; Reinhardt-Rutland, 1987), visible object size (Blakemore & Sutton, 1969), and color (Webster & Mollon, 1991, 1994). However, similar perceptual aftereffects following adaptation are observed across many other research areas, including face perception (e.g., Jones et al., 2010; Leopold, Rhodes, Muller, & Jeffery, 2005; Rhodes et al., 2004; Webster, Kaping, Mizokami, & Duhamel, 2004), audition (e.g., Anstis & Saida, 1985; Ehrenstein, 1994; Grantham & Wightman, 1979), touch (e.g., Gibson & Backlund, 1963; Vogels, Kappers, & Koenderink, 1996), and kinesthesis (e.g., Day & Singer, 1964; Jaffe, 1956; Kohler & Dinnerstein, 1947; Nachmias, 1953)

As discussed previously, Roberts and Summerfield (1981) failed to find significant crossmodal adaptation between auditory and visual speech. However, adaptation to some non-speech information *can* induce perceptual aftereffects across sensory modalities (for a review, see Rosenblum, Dias, et al., 2016). For example, following adaptation to visual horizontal motion, perceivers identify stationary auditory sounds as moving in the opposite direction of the adapted visual motion (Ehrenstein & Reinhardt-Rutland, 1996). Similarly, adaptation to visual motion in depth transfers to an auditory motion aftereffect (perceived changes in intensity), though adaptation to auditory motion in depth does not transfer to a visual motion aftereffect (Kitagawa & Ichihara, 2002). More recently, a bidirectional aftereffect have been observed between visual and tactile motion, such that adaptation to visual motion in one direction transfers to perceived haptic motion in the opposite direction, and vice versa (Konkle, Wang, Hayward, & Moore, 2009). Bidirectional crossmodal aftereffects have also been observed for the identification of faces by touch and vision (Matsumiya, 2013) and for the perceived timing rate of auditory and visual stimuli (Levitan, Ban, Stiles, & Shimojo, 2015)

Also, unlike speech, *redundant* information provided across sensory modalities can enhance non-speech aftereffects. For example, adaptation to haptic motion with visual motion can produce a visual motion aftereffect that is stronger than the aftereffect produced by visual adaptation alone (Matsumiya & Shioiri, 2008). Similarly, simultaneous adaptation to visual and auditory motion, both in depth (Kitagawa & Ichihara, 2002) and in direction (Vroomen & de Gelder, 2003), can produce auditory motion aftereffects that are stronger than the aftereffects produced by auditory adaptation alone.

Purpose of the current investigations

As stated earlier, evidence from much multisensory speech research suggests crossmodal influences occur early in the speech process, perhaps even prior to the extraction of phonetic features (for a review, see Rosenblum, Dorsi, et al., 2016). However, the failure to find audio-visual crossmodal and bimodal influences in selective speech adaptation (e.g., Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994) suggests that earlier sensory-specific processes may affect speech perception prior to the integration and extraction of modality-neutral information across sensory modalities (e.g., Massaro, 1987; Samuel & Lieblich, 2014).

Failing to find auditory-visual crossmodal influences in selective speech adaptation is odd considering such crossmodal adaptation has been observed in non-speech domains (for a review, see Rosenblum, Dias, et al., 2016). It is especially odd considering Cooper and colleagues found selective speech adaptation between auditory speech perception and (the kinesthetic and tactile sensations associated with) speech production (Cooper, 1974; Cooper et al., 1976; Cooper et al., 1975).

It is important to recognize, however, that Roberts and Summerfield (1981) provide the only published study explicitly investigating crossmodal audio-visual speech adaptation. They also provide the only published study investigating audio-visual bimodal enhancement of speech adaptation. That Roberts and Summerfield (1981) provide the only investigation to date for such auditory-visual influences in speech adaptation suggests that more investigation is needed.

The goal of the current series of investigations is to examine whether any crossmodal influences between auditory and visual speech exist in selective speech adaptation. The investigations share a common theme of understanding what information is important for selective speech adaptation; whether that information depends on sensory overlap between adaptors and test stimuli, or whether that information can take a modality-neutral phonetic form.

Chapter 1: Influences of selective speech adaptation on perception of audiovisual speech

As previously discussed, Roberts and Summerfield (1981) observed changes in auditory speech identification following adaptation to a bimodal auditory-/ba/-visual-/ga/ adaptor, which is typically perceived as /da/ (e.g., McGurk & MacDonald, 1976). However, the changes in auditory speech identification reflected adaptation to the auditory component of the audiovisual adaptor, not the audiovisual-/da/ percept (see also Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014).

Chapter 1 proposes that the use of such McGurk-like audiovisual adaptors may undercut observation of crossmodal influences in selective speech adaptation. Previously, Rosenblum and Saldaña (1992) found that when comparing incongruent audio-visual combinations (auditory-/ba/-visual-/va/, typically perceived as /va/) and congruent audio-visual combinations (auditory-/va/-visual-/va/) to auditory-only examples of the same phonetic information (auditory-/va/), perceivers are much more likely to identify the congruent audio-visual combinations as more similar to the auditory-only examples. The results suggest that phonetic percepts derived from incongruent audio-visual combinations are more sensitive (ambiguous) perceptual objects. The sensitive nature of these phonetic percepts may suggest they serve as poor *adaptors* in selective speech adaptation (e.g., Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014). However, the sensitive nature of these phonetic percepts may also suggest they can serve as *test-stimuli* that are more susceptible to crossmodal influence in selective speech adaptation.

Chapter 1 (previously published as Dias, Cook, & Rosenblum, 2016), investigates whether adaptation to phonetic information provided across sensory modalities can influence perception of integrated audio-visual phonetic information. Changes in the identification of test-stimuli along an *audiovisual* continuum were measured following adaptation to auditory and visual syllables. The test-continuum consisted of nine audio-/ba/-visual-/va/ stimuli, ranging in visibility of the articulating mouth. When visibility of the mouth was unobstructed, the auditory-/ba/-visual-/va/ stimulus was identified as /va/ 93.7% of the time (e.g., McGurk & MacDonald, 1976). When visibility of the mouth was occluded, the auditory-/ba/-visual-/va/ stimulus was identified as /ba/. Tokens in the middle of the continuum were identified half of the time as /va/ and half of the time as /ba/.

I hypothesized that if crossmodal influences exist in selective speech adaptation, then identification of audiovisual speech should change following adaptation to any adaptor that shared salient phonetic information with the audiovisual test-stimuli (auditory-/va/, auditory-/ba/, visual-/va/, visual-/ba/, and audiovisual-/va/). However, if selective speech adaptation depends on sense-specific spectrotemporal overlap between adaptors and test-stimuli, then changes in audiovisual speech perception should only follow adaptation to adaptors that share salient *sense-specific* phonetic information with the visual component of the test-stimuli (visual-/va/ and audiovisual-/va/).

Chapter 2: Selective adaptation of crossmodal speech information: A case of small but consistent effects

Roberts and Summerfield (1981) failed to find a significant change in auditory speech identification following visual speech adaptation. As previously discussed, their failure to observe crossmodal influence in selective speech adaptation is inconsistent with evidence from other speech domains and evidence from non-speech domains demonstrating crossmodal adaptation between audition and vision. However, a close examination of the data reported by Roberts and Summerfield (1981) may suggest subtle (non-significant) crossmodal adaptation that warrants further investigation.

Chapter 2 employed meta-analyses and large sample sizes to examine whether subtle crossmodal influences in selective speech adaptation exist. I hypothesized that if selective speech adaptation is influenced by modality-neutral information, then changes in auditory speech identification should follow adaptation to visual speech, and changes in visual speech identification should follow adaptation to auditory speech. However, if selective speech adaptation depends on shared sense-specific spectrotemporal overlap between adaptors and test-stimuli, then changes in auditory speech identification should only follow adaptation to auditory speech, and changes in visual speech identification should only follow adaptation to visual speech.

Chapter 3: Audio-visual selective speech adaptation does not exhibit a bimodal advantage

Chapter 3 evaluates whether combined audio-visual information in speech adaptors can induce greater perceptual changes than unimodal adaptors. In one condition, identification of ambiguous auditory segments was measured before and after auditory and audiovisual speech adaptation. In another condition, identification of ambiguous visual segments was measured before and after visual and audiovisual speech adaptation. I hypothesized that if subtle crossmodal influences exist in selective speech adaptation, then changes in speech identification should be greater following adaptation to bimodal adaptors than following adaptation to unimodal adaptors.

The dissertation will conclude with a discussion of the theories that may account for the observations across the three chapters. Theoretical accounts explaining the current results will also need to account for crossmodal influences observed in other speech domains (for a review, see Rosenblum, Dorsi, et al., 2016) and for the influences of lexical knowledge on selective speech adaptation (e.g., Samuel, 1997, 2001; Samuel & Lieblich, 2014).

References

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance, 11*(3), 257-271.

Anstis, S., Verstraten, F. A. J., & Mather, G. (1998). The motion aftereffect. *Trends in Cognitive Science, 2*(3), 111-117.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92*(2), 339-355.

Baart, M., & Vroomen, J. (2010). Do you see what you are hearing? Cross-modal effects of speech sounds on lipreading. *Neuroscience Letters, 471*, 100-103.

Bernstein, L. E., & Liebenthal, E. (2014). Neural pathways for visual speech perception. *Frontiers in Neuroscience, 8*(386). doi:10.3389/fnins.2014.00386

Blakemore, C., & Sutton, P. (1969). Size adaptation: A new aftereffect. *Science, 166*, 245-247.

Callan, D. E., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: A single-sweep EEG case study. *Cognitive Brain Research, 10*, 349-353.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., . . . David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*, 593-596.

Campbell, C. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society B, 363*, 1001-1010.

Carleson, V. R. (1962). Adaptation in the perception of visual velocity. *Journal of Experimental Psychology, 64*(2), 192-197.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology, 113*, 495-506.

Cooper, W. E. (1974). Perceptuomotor adaptation to a speech feature. *Perception & Psychophysics, 16*(2), 229-234.

Cooper, W. E., Billings, D., & Cole, R. A. (1976). Articulatory effects on speech perception: A second report. *Journal of Phonetics, 4*(3), 219-232.

Cooper, W. E., Blumstein, S. E., & Nigro, G. (1975). Articulatory effects on speech perception: A preliminary report. *Journal of Phonetics, 3,* 87-98.

Day, R. H., & Singer, G. (1964). Spatial aftereffects within and between kenesthesis and vision. *Journal of Experimental Psychology, 68*(4), 337-343.

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. doi:10.1016/j.wocn.2016.02.004

Dias, J. W., Cook, T. C., & Rosenblum, L. D. (in press). The McGurk effect and the primacy of multisensory perception. In A. G. Shapiro & D. Todorovic (Eds.), *Oxford Compendium of Visual Illusions*: Oxford University Press.

Dias, J. W., & Rosenblum, L. D. (2011). Visual influences on interactive speech alignment. *Perception, 40*, 1457-1466.

Dias, J. W., & Rosenblum, L. D. (2015). Visibility of speech articulation enhances auditory phonetic convergence. *Attention*, *Perception, & Psychophysics*, *77*(6). doi:10.3758/s13414-015-0982-6

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology, 1*(2), 121-144.

Diehl, R. L., Kluender, K. R., & Parker, E. M. (1985). Are selective adaptation and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception and Performance*, *11*(2), 209-220.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Carnegie Mellon University Research Showcase*. Retrieved from http://repository.cmu.edu/psychology/155

Ehrenstein, W. H. (1994). Auditory aftereffects following simulated motion produced by varying interaural intensity or time. *Perception, 23*(10), 1249-1259.

Ehrenstein, W. H., & Reinhardt-Rutland, A. H. (1996). A cross-modal aftereffect: Auditory displacement following adaptation to visual motion. *Perception and Motor Skills*, *82*, 23-26.

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

Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders, 40*(4), 481-492.

Fowler, C. A. (2004). Speech as a supramodal or amodal phenomenon. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processing* (pp. 189-202). Cambridge, MA: MIT Press.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17*(3), 816-828.

Ganong, W. F. (1978). The selective adaptation effects of burst-cued stops. *Perception & Psychophysics, 24*(1), 71-83.

Gibson, J. J., & Backlund, F. A. (1963). An after-effect in haptic space perception. *The Quarterly Journal of Experimental Psychology, 15*(3), 145-154.

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature, 426*, 502-504.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251-279.

Goldstein, L. M., & Lackner, J. R. (1972). Alterations of the phonetic coding of speech sounds during repetition. *Cognition, 3*, 279-297.

Grantham, D. W., & Wightman, F. L. (1979). Auditory motion aftereffects. *Perception & Psychophysics, 26*(5), 403-408.

Green, K. P. (1996). *Studies of the McGurk effect: Implications for theories of speech perception*. Paper presented at the Fourth International Conference on Spoken Language, Philadelphia, PA.

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *PNAS, 106*(4), 1245-1248.

Jaffe, R. (1956). The Influence of Visual Stimulation on Kinesthetic Figural After-Effects. *The American Journal of Psychology, 69*(1), 70-75. doi:10.2307/1418116

Jones, B. C., Feinberg, D. R., Bestelmeyer, P. E. G., DeBruine, L. M., & Little, A. C. (2010). Adaptation to different mouth shapes influences visual perception of ambiguous lip speech. *Psychonomic Bulletin & Review, 17*(4), 522-528.

Kat, D., & Samuel, A. G. (1984). More adaptation of speech by nonspeech. *Journal of Experimental Psychology: Human Perception and Performance, 10*(4), 512-525.

Kim, J., Davis, C., & Krins, P. (2004). Amodal processing of visual speech as revealed by priming. *Cognition, 93*(1), B39-B47. doi:http://dx.doi.org/10.1016/j.cognition.2003.11.003

Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature, 416*, 172-174.

Kohler, W., & Dinnerstein, D. (1947). Figural after-effects in kinesthesis. *Miscellanea Psychologica Albert Michotte*, 196-220.

Konkle, T., Wang, Q., Hayward, V., & Moore, C. I. (2009). Motion aftereffects transfer between touch and vision. *Current Biology, 19*, 1-6.

Lachs, L., & Pisoni, D. B. (2004a). Cross-modal source information and spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 30*(2), 378-396.

Lachs, L., & Pisoni, D. B. (2004b). Crossmodal source identification in speech perception. *Ecological Psychology, 16*(3), 159-187.

Lachs, L., & Pisoni, D. B. (2004c). Specification of cross-modal source information in isolated kinematic displays of speech. *Journal of the Acoustical Society of America, 116*(1), 507-518.

Leopold, D. A., Rhodes, G., Muller, K. M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society, Biology, 272*, 897-904.

Levitan, C. A., Ban, Y.-H. A., Stiles, N. R. B., & Shimojo, S. (2015). Rate perception adapts across the senses: evidence for a unified timing mechanism. *Sci. Rep., 5*. doi:10.1038/srep08857

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24*(3), 253-257.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (2015). Speech perception. In J. D. Wright (Ed.), *International Encyclopedia of Social & Behavioral Sciences* (2nd ed., Vol. 23, pp. 235-242). Oxford: Elsevier.

Matsumiya, K. (2013). Seeing a haptically explored face: Visual facial-expression aftereffect from haptic adaptation to a face. *Psychological Science, in press*.

Matsumiya, K., & Shioiri, S. (2008). Haptic movements enhance visual motion aftereffect [Abstract]. *Journal of Vision, 8*(6), 172.

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

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics, 72*(6), 1614-1625.

Musacchia, G., Sams, M., Nicol, T., & Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Experimental Brain Research, 168*(1-2), 1-10.

Nachmias, J. (1953). Figural After-Effects in Kinesthetic Space. *The American Journal of Psychology, 66*(4), 609-612. doi:10.2307/1418958

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research, 71,* 4-12.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*(4), 2382-2393.

Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I. P., Mottonen, R., Tarkiainen, A., & Sams, M. (2005). Primary auditory cortex activation by visual speech: An fMRI study at 3T. *Auditory and Vestibular Systems, 16*(2), 125-128.

Reinhardt-Rutland, A. H. (1987). Aftereffect of visual movement-the role of relative movement: A review. *Current Psychological Research & Reviews, 6*(4), 275-288.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lip-Reading* (pp. 97-113). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Remez, R. E. (1980). Susceptibility of a stop consonant to adaptation on a speech-nonspeech continuum: Further evidence against feature detectors in speech perception. *Perception & Psychophysics, 27*(1), 17-23.

Rhodes, G., Jeffery, L., Watson, T. L., Jaquet, E., Winkler, C., & Clifford, C. W. G. (2004). Orientation-contingent face aftereffects and implications for face-coding mechanisms. *Current Biology, 14,* 2119-2123.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics, 30*(4), 309-314.

Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. Pisoni & R. Remez (Eds.), *Handbook of Speech Perception* (pp. 51-78). Malden: Blackwell.

Rosenblum, L. D., Dias, J. W., & Dorsi, J. J. (2016). The supramodal brain: implications for auditory perception. *Journal of Cognitive Psychology, 2016*. doi:10.1080/20445911.2016.1181691

Rosenblum, L. D., Dorsi, J. J., & Dias, J. W. (2016). The impact and status of Carol Fowler's supramodal theory of multisensory speech perception. *Ecological Psychology*.

Rosenblum, L. D., Miller, R. M., & Sanchez, K. (2007). Lip-read me now, hear me better later: Cross-modal transfer of talker-familiarity effects. *Psychological Science, 18*(5), 392-396.

Rosenblum, L. D., & Saldaña, H. M. (1992). Discrimination tests of visually influenced syllables. *Perception and Psychophysics, 52*(4), 461-473.

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 22*(2), 318-331.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17*, 1147-1153.

Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *Journal of the Acoustical Society of America, 95*(6), 3658-3661.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127*, 141-145.

Sams, M., Mottonen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research, 23*, 429-435.

Samuel, A. G. (1981). Phonemic Restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*(4), 474-494.

Samuel, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Langauge, 26*, 36-56.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32*, 97-127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*(4), 348-351.

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), 1479-1490.

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.

Sanchez, K., Dias, J. W., & Rosenblum, L. D. (2013). Experience with a talker can transfer across modalities to facilitate lipreading. *Attention, Perception, & Psychophysics, 75*(7), 1359-1365.

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*(4), 661-699.

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

Vogels, I. M. L. C., Kappers, A. M. L., & Koenderink, J. J. (1996). Haptic aftereffect of curved surfaces. *Perception, 25*, 109-119.

Vroomen, J., & de Gelder, B. (2003). Visual motion influences the contingent auditory motion aftereffect. *Psychological Science, 14*(4), 357-361.

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature, 428*, 557-561.

Webster, M. A., & Mollon, J. D. (1991). Changes in colour appearance following post-receptoral adaptation. *Nature, 349*, 235-238.

Webster, M. A., & Mollon, J. D. (1994). The influence of contrast adaptation on color appearance. *Vision Research, 34*(15), 1993-2020.

Chapter 1

Influences of Selective Adaptation on Perception of Audiovisual Speech

The materials in this chapter have previously been published as: 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. doi:10.1016/j.wocn.2016.02.004. These materials have been reproduced for the purposes of this dissertation with permission from the publisher.

Speech is a multimodal phenomenon (for a review, see Rosenblum (2008)). Visual speech information can improve identification of auditory speech presented in difficult listening conditions (e.g., Erber, 1975; Remez, Fellowes, Pisoni, Goh, & Rubin, 1998; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954), and enhance intelligibility of speech that conveys complicated content (e.g., Arnold & Hill, 2001; Reisberg, McLean, & Goldfield, 1987). Perceivers will subtly imitate the speech characteristics of a perceived talker after listening to (e.g., Goldinger, 1998; Pardo, 2006) and lipreading (Miller, Sanchez, & Rosenblum, 2010) the speech of that talker, demonstrating how heard and seen speech modulate speech production.

The most striking demonstrations of the multimodal nature of speech perception are phenomena where perception of an acoustic speech signal is modified by conflicting information provided by another sensory modality. For example, the McGurk effect (McGurk & MacDonald, 1976) demonstrates how perception of auditory speech can be modulated by incongruent visual speech information. An auditory-/ba/ presented in synchrony with a visible articulation of “va” (visual-/va/) is typically perceived as “va” (e.g., Rosenblum & Saldaña, 1992). McGurk-like effects have been demonstrated when auditory speech information is paired with conflicting articulatory information provided by other sensory modalities. For example, conflicting kinesthetic (e.g., Ito, Tiede, & Ostry, 2009; Sams, Mottonen, & Sihvonen, 2005) and haptic information (e.g., Fowler & Dekle, 1991; Gick & Derrick, 2009) can also influence how auditory speech is perceived. The illusory percepts resulting from the conflicting sensory information are often described as a resolution of the shared articulatory information available across the conflicting sensory inputs. As such, the information across sensory modalities integrates to produce a unified percept that shares information with the conflicting sensory inputs (e.g., McGurk & MacDonald, 1976).

A question in the speech literature regards at what point in speech processing cross-sensory information completely integrates. While some theories propose that information across sensory modalities is completely integrated early in the speech process (for reviews, see Fowler (2004), Rosenblum (2008)), other theories propose that cross-sensory information is integrated only after some initial processing of sensory information (for reviews, see Bernstein, Auer, and Moore (2004) and Massaro (1987)). Selective adaptation in speech perception provides a behavioral paradigm for investigating low-level sensory influences on phonetic perception (for a review, see Vroomen & Baart, 2012). In the following investigation, we explore whether auditory and visual speech information fully integrate by the time information reaches the early level at which selective adaptation is thought to occur.

Selective Adaptation in Speech Perception

Previous research has used the ability of perceivers to selectively adapt to perceived speech as a metric for investigating the nature of the speech recognition mechanism. Traditionally, selective adaptation in speech is evaluated by testing the effects of listening to repeated presentations of specific syllable adaptors on perception of syllable tokens along a test continuum, which ranges from one phonetic category to another. Following adaptation, perceivers can exhibit a boundary shift between perceived phonetic categories. For example, Eimas and Corbit (1973) originally examined how adaptation to repeated presentations of auditory /ba/ or /pa/ syllables could shift the perceived phonetic boundary along a 14-item auditory /ba/-/pa/ continuum. Hearing a repeated /ba/ resulted in more items along the continuum identified as /pa/ (a phonetic boundary shift towards /ba/). Conversely, adaptation to /pa/ resulted in more items along the continuum identified as /ba/ (a phonetic boundary shift towards /pa/).

The original explanation for selective adaptation is that the repetition of a syllable stimulus serves to fatigue a "linguistic feature detector"; a hypothetical mechanism thought to be sensitive to specific featural, or phonetic, characteristics of speech sounds (e.g. Eimas, Cooper, & Corbit, 1973; Eimas & Corbit, 1973). The result is a deficit in subsequent sensitivity to that phonetic characteristic. For example, returning to the /ba/-/pa/ experiment described above, the perceptual shifts following adaptation to /ba/ or /pa/ occur because each adaptor fatigues perception of their respective voice-onset-time (VOT) characteristic. Thus, adaptation to /ba/ fatigues perception of short VOTs, resulting in more items along the /ba/ to /pa/ continuum perceived as having longer VOTs, consistent with a /pa/ percept. Conversely, adaptation to /pa/ fatigues perception of long VOTs, resulting in more items along the /ba/ to /pa/ continuum perceived as having shorter VOTs, consistent with a /ba/ perception. To emphasize the point, Eimas and Corbit (1973) demonstrated how adaptation to stimuli sharing VOT characteristics with the test continua could shift perceived phonetic boundaries even in other phonemes with similar VOT features. For example, adaptation to audio-/da/ could shift phonetic categories along a /ba/-to-/pa/ continuum in a way similar to audio-/ba/.

One question about selective adaptation in speech is whether the adaptation effects are purely auditory in nature; dependent on shared acoustic information between an adaptor and test stimulus. Auditory accounts are supported by findings illustrating that adaptation effects are greater when there is more spectral overlap between the adaptor and test stimuli (e.g., Ganong, 1978). Other evidence showing that perception of auditory speech can be modulated by adaptation to non-speech acoustic information (e.g., white noise) further supports auditory accounts (e.g. Kat & Samuel, 1984).

However, there is also evidence that visual speech adaptors can shift perception of continua involving visual speech components. For example, Jones, Feinberg, Bestelmeyer, DeBruine, and Little (2010) found that adapting perceivers to still images of mouth shapes articulating /m/ or /u/ speech sounds could shift perceptual boundaries along an /m/-to-/u/ continuum of still-face images; adaptation to /m/ resulting in more continuum items being identified as /u/; and adaptation to /u/ resulting in more items identified as /m/. These visual adaptation effects occurred even when the adaptor image involved a model different from that of the test-continuum images. This finding could suggest that perceivers can adapt to the general gestural state of a face image, as opposed to idiosyncratic characteristics associated with a specific talker's face.

Baart and Vroomen (2010) found similar results for videos of faces dynamically articulating speech sounds. The visual test continuum used in the Baart and Vroomen (2010) study was created by overlaying visual utterances of /onso/ and /omso/ while adjusting the opacity of the overlaid images. The final continuum subtly transitioned from low opacity of /onso/ and high opacity of /omso/ at the /onso/-end of the continuum to low opacity of /omso/ and high opacity of /onso/ at the /omso/-end of the continuum. Following repeated exposure to an audiovisual-recorded model uttering /onso/, perceivers identified more ambiguous visual stimuli along an /onso/-to-/omso/ continuum as /omso/. Conversely, perceivers identified more ambiguous visual stimuli as /onso/ following repeated exposure to audiovisual /omso/. The results further suggest that selective adaptation of visual speech information can influence subsequent perception of visual speech.

The evidence demonstrating selective adaptation effects for visual speech information suggests that selective adaptation in speech is not an auditory-only phenomenon. This could mean that selective adaptation in speech depends on common, amodal phonetic information shared between the adaptor and test stimuli. If such a premise is true, then it would suggest that illusory phonetic information integrated across sensory modalities can induce adaptation effects.

Two studies have explicitly investigated this question, measuring changes in auditory-phonetic perception following adaptation to audiovisual discrepant adaptors that produce integrated phonetic percepts (e.g., McGurk & MacDonald, 1976). For example, Roberts and Summerfield (1981) found that adaptation to an audio-visual discrepant stimulus (i.e., audio-/bɛ/-visual-/gɛ/, often perceived as "dɛ") results in a phonetic boundary shift towards /bɛ/ along a /bɛ/ to /dɛ/ auditory test-continuum. In other words, perceivers demonstrated adaptation to the auditory component of the audiovisual adaptor, despite often reporting a percept influenced by the visual component. Further, adaptation to visual-only representations of /bɛ/ or /dɛ/ produced nonsignificant shifts in perceived phonetic categories along the auditory test-continuum. Saldaña and Rosenblum (1994) demonstrated similar results using an auditory-/ba/-visual-/va/ adaptor, which has typically been found to produce a visually-influenced percept (e.g., 'heard' "va") more regularly than the audio-/bɛ/-visual-/gɛ/ stimulus. In fact, Saldaña and Rosenblum (1994) found that when presented with auditory-/ba/-visual-/va/, perceivers reported 'hearing' "va" 99% of the time. Still, adaptation to an auditory-/ba/-visual-/va/ stimulus shifted phonetic category boundaries along an auditory /ba/ to /va/ continuum toward /ba/; i.e., in the direction of the auditory component of the audiovisual adaptor, similar to the observations of Roberts and Summerfield (1981).

The results of Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994) demonstrate how an adaptor with discrepant audio-visual components shifts phonetic boundaries along an auditory continuum based on the shared auditory information between the adaptor and the test continuum, and not the integrated phonetic information perceived in an audio-visual adaptor. In fact, even when the discrepant audio-visual streams form a lexical percept (e.g., auditory-/armabillo/-visual-/armagillo/ perceived as the real word “armadillo”), adaptation will still fail to produce a measureable shift in auditory speech perception based on the integrated audiovisual percept (Samuel & Lieblich, 2014).

From these studies, it does not appear to be the case that integrated audiovisual information in the adaptor modulates phonetic perception in auditory test-stimuli. This may suggest that auditory and visual speech information are not completely integrated at the level of selective adaptation. However, there may be some problems associated with using adaptor stimuli consisting of incongruent audio-visual speech information to test for integrated phonetic influences in selective adaptation.

There is evidence that percepts based on incongruent audio-visual information (e.g., McGurk & MacDonald, 1976) do not exhibit the same quality of phonetic information compared to that from congruent audio-visual information. For example, audio-visual congruent stimuli (e.g., audio-/va/-visual-/fa/) are preferentially chosen over audio-visual incongruent stimuli (e.g., audio-/ba/-visual-/fa/) as better matches to audio-only phonetic utterances (e.g., audio-/va/), even when the audio-visual incongruent stimulus is perceived as an integrated percept (e.g., heard as “va”) 96% of the time (Rosenblum & Saldaña, 1992). In fact, data across the literature investigating the McGurk effect illustrates how integrated percepts derived from incongruent audio-visual streams can be highly variable. Different audio-visual combinations produce different phonetic percepts at varying rates, and a single audio-visual incongruent stimulus can be perceived as multiple phonetic percepts (e.g., MacDonald & McGurk, 1978; Mallick, Magnotti, & Beauchamp, 2015). Recent evidence even suggests that there is a great deal of variability in how individual perceivers integrate incongruent audio-visual speech information (Mallick et al. 2015). Thus, it could be that audio-visual incongruent stimuli produce more sensitive perceptual objects than percepts derived from unimodal or audio-visual congruent stimuli. The sensitivity of these integrated percepts may qualify them as poor adaptors within a selective adaptation framework (e.g., Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994).

However, the sensitive nature of audio-visual integrated speech percepts may also render them more susceptible to crossmodal influence following adaptation to clear unimodal speech adaptors. In other words, though adaptation to audio-visual integrated speech percepts fails to change auditory speech perception, adaptation to auditory (or visual) speech may change perception of audio-visual integrated percepts.

The Current Investigation

Instead of evaluating how adaptation to audio-visual speech modulates perception of auditory speech (Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994), the goal of the current investigation is to determine whether unimodal auditory or visual speech adaptors can modulate perception of test items comprised of audiovisual speech. The adaptors we employ share varying amounts of cross-sensory and sensory-specific phonetic information with an audiovisual speech continuum constructed for this investigation. The degree to which adapted phonetic information modulates perception of audiovisual speech may depend on the sensory overlap between the adaptor and the test stimuli. This result would be consistent with the findings of Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994). Such an observation would suggest that either the audio and visual streams do not integrate at the level of selective adaptation or, if they do, that the integration is weak or incomplete (so that the separate sensory components of the audiovisual stimulus can still be influenced). However, it may be the case that adaptation to phonetic information available across auditory and visual speech will change perception of integrated audio-visual percepts. These results would suggest that auditory and visual speech information integrate by the time information reaches the level of selective adaptation, at least to a degree that the integrated information is susceptible to crossmodal influence.

We constructed an audiovisual speech test continuum by systematically manipulating the amount of salient visual information available to influence the syllable percept. For our target tokens, we chose an auditory-/ba/-visual-/va/ McGurk stimulus, which is known to be an especially strong visually-influenced combination, with subjects reportedly 'hearing' the syllable as "va" up to 99% of the time (e.g. Saldaña & Rosenblum, 1994). It was important for the visually-influenced syllable to be compelling in order to examine the relative influence of crossmodal-phonetic and sensory-specific adaptation on perception of target-stimuli.

We chose to create our audiovisual-token continuum so that it ranged from a strong visually-influenced "va" percept, to a strong "ba" percept – when the visual component provides minimal articulatory information. To achieve this, the salience of the visual-/va/ component of our audiovisual tokens was modulated using a Gaussian blur technique. This technique has been used previously to create a perceptual continuum of audiovisual tokens: Thomas and Jordan (2002) reported that the strength of the McGurk effect (i.e. the probability of perceiving an auditory-/ba/-visual-/ga/ stimulus as “da”) decreased as the visual stimulus is masked by Gaussian blurring. Greater Gaussian blurring can mask enough of the visual information to nearly eliminate the visual influence on perception of the auditory speech sound (“ba”), with several magnitudes of moderate blurring demonstrating more ambiguous audiovisual percepts. The most ambiguous tokens in their continuum were perceived half of the time as /da/ and half of the time as /ba/. Our audiovisual /va/-to-/ba/ continuum was constructed in an analogous way so that it ranged from a strong unambiguous “va” percept, through more ambiguous tokens, ending with a strong unambiguous “ba” percept. This allowed us to then test how different adaptors might shift perception of the more ambiguous mid-continuum audiovisual tokens.

Adaptation to four different uni-sensory stimuli and one bimodal stimulus was tested to determine the influence of adaptation to shared cross-sensory phonetic and sensory-specific phonetic information on perception of the audiovisual test continuum (see Table 1.1 for a summary). We define cross-sensory phonetic information as information available across sensory modalities and sensory-specific phonetic information as information available only within a specific sensory modality.

Auditory-/va/ served as our critical test-adaptor. Auditory-/va/ shares cross-sensory phonetic information with the visual /va/ component of the audiovisual test-continuum. It also shares cross-sensory phonetic information with the part of the audiovisual test-continuum that produces integrated audio-visual percepts (i.e., audio-/ba/-visual-/va/, heard as “va”). However, because the auditory component of the audiovisual teststimuli is always an unambiguous /ba/, the auditory-/va/ adaptor does not share sensory-specific phonetic information with the (initial segment of the) test stimuli. If selective adaptation can modulate perception of integrated audio-visual phonetic information, then the perceived phonetic boundary between /ba/ and /va/ should shift towards /va/ following adaptation to auditory-/va/ (more ‘ba’ responses will be observed). If, on the other hand, the influence of selective adaptation depends on shared sensory-specific information between the adaptor and test stimuli, then adaptation to auditory-/va/ should not produce a significant phonetic boundary shift.

A visual-/va/ adaptor was also tested. The visual-/va/ adaptor shares sensory-specific (visual) phonetic information with the test continuum, which varies in the clarity of visual-/va/ information. However, the visual-/va/ adaptor also shares some amount of crosssensory phonetic information with the integrated audio-visual percept of our audiovisual test-continuum. This adaptor primarily tests whether adaptation to visual information can modulate processing of visual information in the audio-visual test stimuli, similar to how adaptation to visual information has previously been found to modulate phonetic perception along visual-speech continua (e.g., Jones et al., 2010). If selective adaptation can modulate perception of integrated audio-visual phonetic information, then the perceived phonetic boundary between /ba/ and /va/ should shift towards /va/ following adaptation to visual-/va/. If, on the other hand, selective adaptation can modulate perception of audiovisual speech by influencing shared sensory information between the adaptor and test stimuli, then adaptation to visual-/va/ should still produce a significant phonetic boundary shift towards /va/. In that the prediction is the same whether adaptation is to cross-sensory or sensory-specific information, this adaptor on its own cannot determine the basis of adaptation. However, it can help establish whether our visual adaptor can be influential.

We also tested an audiovisual-/va/ adaptor. This stimulus was comprised of (congruent) auditory-/va/ and visual-/va/ components. Similar to the visual-/va/ adaptor, the audiovisual-/va/ adaptor shares sensory-specific (visual) phonetic information with the test continuum. However, both the auditory and visual components of the audiovisual-/va/ adaptor share some amount of cross-sensory phonetic information with the integrated audio-visual percept of our audiovisual test-continuum. This adaptor primarily tests whether adaptation to congruent audiovisual information can modulate processing of visual information in the audio-visual test stimuli. As previously stated, adaptation to congruent audio-visual information has been found to modulate phonetic perception along visualspeech continua (e.g., Baart & Vroomen, 2010). If selective adaptation can modulate perception of integrated audiovisual phonetic information, then the perceived phonetic boundary between /ba/ and /va/ should shift towards /va/ following adaptation to audiovisual-/va/. If, on the other hand, selective adaptation can only influence shared sensory information between the adaptor and test stimuli, then adaptation to audiovisual-/va/ should still produce a significant phonetic boundary shift towards /va/. However, we made one more prediction based the audiovisual-/va/ adaptor: If selective adaptation, dependent on shared sensory information between the adaptor and test stimuli, can be enhanced by redundant phonetic information provided across sensory modalities, then adaptation to audiovisual-/va/ should produce a greater phonetic boundary shift towards /va/ than the visual-/va/ adaptor.

An auditory-/ba/ adaptor was also tested, which shares cross-sensory phonetic information with the audiovisual test-continuum; as salient visual-/va/ information is obscured, the auditory-/ba/ component has greater influence on the perceived integrated phonetic percept, resulting in more “ba” percepts. Auditory-/ba/ also shares sensory-specific phonetic information with the auditory component of our audiovisual test-continuum. However, the auditory component of our test-continuum is unambiguously /ba/ for all continuum tokens. Recall that adaptation effects are typically observed to modulate perception of only the most ambiguous tokens along a phonetic test-continuum. As such, we do not expect adaptation to auditory-/ba/ to shift the perceived phonetic boundary along the audiovisual test continuum if selective adaptation modulates processing of sensory-specific information shared between the adaptor and test stimuli. We hypothesize that if selective adaptation modulates perception of integrated audiovisual phonetic information, then the perceived phonetic boundary between /ba/ and /va/ should shift towards /ba/ following adaptation to auditory-/ba/. If, on the other hand, selective adaptation modulates perception of audiovisual speech by influencing shared sensory information between the adaptor and test stimuli, then adaptation to audio-/ba/ should not produce a significant phonetic boundary shift.

Finally, we tested a visual-/ba/ adaptor. Similar to auditory-/va/, visual-/ba/ shares cross-sensory phonetic information with the percepts of the audiovisual test continuum, but does not share any sensory-specific information. If selective adaptation modulates perception of integrated audiovisual phonetic information, then the perceived phonemic boundary between /ba/ and /va/ should shift towards /ba/ following adaptation to visual-/ba/. If, on the other hand, selective adaptation can only influence shared sensory-specific phonetic information between the adaptor and test stimuli, then adaptation to visual-/ba/ should not produce a significant phonetic boundary shift.

In sum, if selective adaptation modulates perception of integrated audio-visual speech by influencing processing of crossmodal phonetic information, then perceptual shifts should be observed for all of the adaptors tested (auditory-/va/, visual-/va/, audiovisual-/ va/, auditory-/ba/, and visual-/ba/). Essentially, any adaptor that shares cross-sensory phonetic information with the audiovisual test continuum is expected to have some influence on perception of the integrated audio-visual information in our test continuum. These results would suggest that auditory and visual speech information integrate by the time the information reaches the level of selective adaptation, at least to a degree that it is susceptible to crossmodal influence.

Methods

Participants

Fifty undergraduates, 23 male and 27 female between 18 and 26 years of age (*M*=19.48, *SE*=.233), from the University of California, Riverside undergraduate participant pool participated in partial fulfillment of course credit. All participants were native speakers of English with normal hearing and normal or corrected-to-normal sight.

They were randomly and evenly distributed between five different groups, each adapted to one of the previously described adaptors.

Materials

All audio-video editing was executed using Final Cut Pro 5 software for Mac OSX.

Audiovisual Test Continuum

First, an auditory-/ba/-visual-/va/ McGurk stimulus (perceived as “va”) was created. A male model (age 28, native English speaking, California native) was digitally audio-video recorded uttering /ba/ and /va/ at 30 frames-per-second (fps) at a size of 640 x 480 pixels. The audio component of a /ba/ utterance was digitally extracted and synchronously dubbed onto a video of the model visually articulating /va/. Synchrony of dubbing was achieved by first matching the auditory onset time of the dubbed auditory component with the original auditory component of the audiovisual stimulus, and then making fine-tuned adjustments to correct for any perceptible asynchrony between the auditory and visual components. A pilot study (*N*=30) determined that this audio-/ba/-visual-/va/ McGurk stimulus was perceived as “va” 93.7% of the time (*SE*=2.12%).

The audio-/ba/-visual-/va/ stimulus was then duplicated to make nine copies. The video portion of each copy was then digitally modified by adding varying degrees of Gaussian blurring over the visible speech articulators (Thomas & Jordan, 2002), between the bridge of the nose and the throat, and between the left and right ear, an area of the face found to be important for audiovisual speech perception (e.g., Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998). Across the nine stimuli, the Gaussian blur was set at a radius of 6, 9, 12, 15, 18, 21, 24, 27, and 30 degrees of rotation. Thus, the nine-item test continuum ranged from weak blurring of the visible articulators, preserving the most salient visual information, to strong blurring of the articulators, where little salient visual information was visible (see Fig. 1). As visual information becomes less salient to the audiovisual stimulus, greater perceptual reliance is placed on the auditory component (Thomas & Jordan, 2002). For the current stimuli, the least blurred stimulus (Gaussian radius of 6) is perceived most often as /va/ and the most blurred (Gaussian radius of 30) is perceived most often as /ba/. All test continuum stimuli were 1,800 ms in length.

Audiovisual Foil Stimuli

The same Gaussian blurring procedure was applied to an audio-visually congruent /ba/ stimulus (audio-/ba/-visual-/ba/), and to an audio-visually congruent /va/ stimulus (audio-/va/-visual-/va/) to be used as foils in a phonetic identification task (e.g., MacDonald & McGurk, 1978). The auditory components of these stimuli were dubbed onto their congruent visual components following the same procedures used for dubbing the audiovisual test-stimuli. The resulting nine audiovisual-/ba/ and nine audiovisual-/va/ stimuli were all 1,800 ms in length, the same length as the test-continuum stimuli. These stimuli were included to foil participants who might otherwise determine that all test-stimuli were the same (either all /va/ or /ba/, depending on whether they strategize with the illusory percept or the unambiguous auditory component of the audiovisual stimuli) (e.g., MacDonald & McGurk, 1978).

Adaptors

The adaptor stimuli were created from the recordings used for the test stimuli. The periods of silence before and after spoken utterances in the test stimuli were edited out of the adaptor stimuli, making them shorter in length (1100 ms). By reducing their length to contain just the available visible and/or auditory speech information within the token, the adaptor stimuli could be presented more often over a shorter period of time during adaptation (described below), yet the stimuli were long enough to contain all visible articulatory information associated with the adapting utterance.

Auditory-/va/. The auditory component of the original audio-video recorded /va/ utterance was digitally extracted and used independently as an adaptor.

Visual-/va/. The visual component of the original audio-video recorded /va/ utterance was digitally extracted, digitized at 30 fps at a size of 640 x 480 pixels, and used as a visual adaptor.

Audiovisual-/va/. The audiovisual-/va/ adaptor was taken from an original audio-video recorded utterance of the male model uttering /va/.

Auditory-/ba/. The audio component of an audio-video recorded /ba/ utterance was digitally extracted and used independently as an adaptor.

Visual-/ba/. The visual component of the original audio-video recorded /ba/ utterance was digitally extracted, digitized at 30 fps at a size of 640 x 480 pixels, and used as a visual adaptor.

Procedure

Baseline Task

Prior to adaptation, baseline phonetic category boundaries were measured using a phonetic identification task. For each trial, an audiovisual stimulus was presented over a computer monitor (24 in ViewSonic VX2450 at 60 Hz and 1920 x 1080 resolution) and headphones (Sony MDR-V600 headphones adjusted to 70 dB SPL) and the participant then identified the token as producing a "ba" or "va" sound. As with previous McGurk studies, participants were instructed to attend to the visual information presented, but to base their judgments on what they heard the speaker say (e.g., MacDonald & McGurk, 1978; McGurk & MacDonald, 1976).

During the baseline task, the nine audio-/ba/-visual-/va/ (A-/ba/-V-/va/) critical test stimuli were presented along with the nine audiovisual-/ba/ (AV-/ba/) and nine audiovisual-/va/ (AV-/va/) foil tokens. Stimuli were presented randomly, but controlled to ensure that one A-/ba/-V-/va/, one AV-/ba/, and one AV-/va/ stimulus was presented every three trials. Each stimulus was presented 5 times over the course of 135 trials ([9 (A-/ba/-V-/va/)+9 (AV-/ba/) +9 (AV-/va/)] x 5 presentations each=135 trials).

Adaptation Task

Upon completion of the baseline task, subjects participated in the critical adaptation task. The adaptation technique used by Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994) was employed for the current experiment, with modifications made to accommodate inclusion of foil trials. Participants were exposed to an initial adaptation phase consisting of 50 exposures to one of the previously described adaptors (100 ms ISI). As with the previous experiments (Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994), this initial adaptation phase was employed to build-up adaptation to the adapted speech information. After this initial adaptation phase, participants underwent 45 additional adaptation cycles. Each cycle consisted of 50 exposures to the adaptor, followed by three speech identification trials. Of the three identification trials presented in each cycle, two were audiovisual foil trials (an AV-/ba/ and an AV-/va/) and one was an audiovisual test trial (audio-/ba/-visual-/va/), presented randomly. Over the course of the 45 cycles, participants completed 135 speech identification trials, with the same stimulus-breakdown as the 135-trial baseline: 45 AV-/ba/ foil trials (9-item continuum, each item presented 5 times), 45 AV-/va/ foil trials (9-item continuum, each item presented 5 times), and 45 A-/ba/-V-/va/ test trials (9-item test-continuum, each continuum item presented 5 times).

Five participant groups were designated based on the adaptor used during the adaptation phase; The audio-/va/, visual-/va/, audiovisual-/va/, audio-/ba/, and visual-/ba/ adaptors were tested between groups.

Results

For tokens of the critical auditory-/ba/-visual-/va/ continuum, participant responses were coded as the proportion of times each of the nine items along the test continuum were identified as /ba/ (see Fig. 2). Similar to previous studies (e.g. Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994), cumulative normal ogives were fitted for the identification performance of each participant prior to and post adaptation, employing the method of probits (Finney, 1971). The number of the hypothetical test stimulus corresponding to the 50% point for each participant's function provided a measure of where the phonetic boundary between /ba/ and /va/ was perceived along the test continuum. Comparisons of the phonetic boundary prior to and post adaptation were conducted for each adaptor group to evaluate the magnitude of phonetic boundary shifts following adaptation to each adaptor stimulus (see Table 1.2).

No significant shift in perceived phonetic boundary was observed for those participants adapted to auditory-/va/, auditory-/ba/, or visual-/ba/. Recall that these uni-sensory adaptors each share cross-sensory phonetic information with the audiovisual test-continuum, but do not share any sensory-specific phonetic information that would be expected to shift perception of phonetic category boundaries across the audiovisual test-continuum.

However, a significant phonetic boundary shift ($p<0.05$) was observed for those participants adapted to visual-/va/ and those participants adapted to audiovisual-/va/: Phonetic category boundaries shifted towards /va/ and more test stimuli were identified as /ba/ following adaptation. A 2-within (baseline, adapted) by 2-between (visual-/va/, audiovisual-/va/ group) mixed-design ANOVA revealed that the magnitude of the phonemic boundary shift between participants adapted to visual-/va/ and participants adapted to audiovisual-/va/ did not significantly differ, $F(1,18)=0.902$, $p=.461$, $\eta_p^2=.030$. This result suggests that the redundant phonetic information provided by the auditory component of the audiovisual-/va/ adaptor did not significantly increase the magnitude of the phonemic boundary shift produced by adaptation to visual-/va/.

The visual-/va/ and audiovisual-/va/ adaptors share cross-sensory and sensory-specific phonetic information with the audiovisual test continuum. Finding a significant phonetic boundary shift only for participants adapted to stimuli containing visual-/va/ suggests that adaptation to cross-sensory phonetic information is insufficient to change perception of integrated audio-visual phonetic percepts. The results are consistent with the findings of Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994). Adaptation to sensory-specific phonetic information seems to change perception of integrated audio-visual phonetic percepts by affecting processing of sensory information shared between the adaptor and test-stimuli. Following adaptation to visual-/va/, participants exhibited a decrease in the degree to which visual-/va/ information could influence perception of the auditory-/ba/ component of the audiovisual test-stimuli. As a result, participants appeared to rely more on the auditory component of the audiovisual test stimuli when making phonetic judgments (e.g., Thomas & Jordan, 2002).

Discussion

Audiovisual speech perception is modulated by selective adaptation only when there is sensory-specific phonetic information shared between the adaptor and test stimuli. We found that changes in phonetic category perception along our test continuum of integrated audiovisual percepts (/va/-to-/ba/) resulted only after adaptation to visual speech information salient to the test continuum (i.e., visual-/va/). Though auditory-/va/, auditory-/ba/, and visual-/ba/ all share cross-sensory phonetic information with the continuum of audiovisual stimuli, adaptation to this information failed to produce any significant changes in audiovisual speech perception. Though we had proposed that percepts resolved from incongruent audio-visual information could provide more sensitive test stimuli for examining crossmodal influences at the level of selective adaptation, we instead find that adaptation effects still depend on sensory information shared between adaptors and test stimuli. This lack of crossmodal influence suggests that auditory and visual speech information do not completely integrate by the time information reaches the level of selective adaptation.

These results are consistent with findings suggesting that integrated audio-visual speech fails to induce selective adaptation effects (Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994). As previously discussed, though auditory-/ba/ does share sensory information with the auditory component of the audiovisual test-continuum, no significant shift in phonetic perception was expected as a result of this shared sensory information. Phonetic boundary shifts following selective speech adaptation are classically observed among ambiguous members of a test continuum. However, the auditory component of our test continuum was always unambiguously /ba/.

What Information Does Selective Adaptation Influence?

Failure to find crossmodal influences in audiovisual speech perception at the level of selective adaptation seems in contrast with robust crossmodal influences in other speech research areas (for a review, see Rosenblum (2008)). Behavioral and neurophysiological evidence suggests that crossmodal integration of speech information occurs early in the speech process (Calvert, Campbell, & Brammer, 2000; Campbell, 2008; Green, 1998; Remez, 2005; Rosenblum, 2005; Summerfield, 1987). The robust and automatic nature of the McGurk Effect itself has served as evidence for a speech process that integrates information across sensory modalities at an early stage in processing, perhaps even at the featural level, prior to the extraction of speech segments (for a review, see Dias, Cook, and Rosenblum (In press) and Rosenblum (2008)). For example, behavioral evidence has demonstrated how information pertaining to the vocal aspiration feature that differentiates /b/ from /p/ can be provided across different sensory modalities to modulate perception of an auditory utterance of /b/. This can be demonstrated by how slowing the visible rate of bilabial articulation can change perception of a normal auditory utterance of /b/ to /p/ (Green & Miller, 1985). Similarly, providing the tactile sensation of an air-burst to the hand or neck in conjunction with auditory /b/ can change perception of a /b/ utterance to /p/ (Gick & Derrick, 2009).

Neurophysiological evidence also suggests early integration of crossmodal speech information. Lipreading can modulate activations in auditory cortex (Calvert et al., 1997; Campbell, 2008; Pekkola et al., 2005) and visual speech information can determine cortical activations in auditory cortex over and above those of auditory speech information (e.g., Callan, Callan, Kroos, & VatikiotisBateson, 2001; Colin et al., 2002; Sams et al., 1991). For example, auditory cortex responds differentially to audio-visual congruent (e.g., audiovisual-/pa/) and audio-visual incongruent (e.g., audio-/pa/-visual-/ka/, typically perceived as “ta” or “ka”) speech tokens, even though the auditory component of the two tokens is the same (Sams et al., 1991). In fact, visual speech information can even modulate neural activity in auditory brainstem (Musacchia, Sams, Nicol, & Kraus, 2006).

It is unclear how the current and past evidence for a sensory-specific basis of selective adaptation can be rationalized with the evidence for early integration and crossmodal influences. However, as stated, evidence for sensory-specific adaptation can only support that some of the information remains nonintegrated. It could be that at the (presumed) early level of adaptation, some integration and crossmodal influences do occur, but not such that adaptation can be influenced. This interpretation could rationalize the current and past selective adaptation findings with the compelling evidence that crossmodal influences can occur early (e.g., see Rosenblum (2008), for a review). Alternatively, it could be that selective adaptation for speech occurs at a level earlier than that of feature extraction (e.g., Green, 1998) and auditory brainstem (Musacchia et al. 2006). These are questions that will need to be addressed in the future.

Another question raised from the results of the current investigation regards the form of the information adapted. In the current investigation, we find that adaptation to visual speech information can change perception of audiovisual speech by modulating the influence of visual speech on perception of auditory speech, not by changing perception of integrated audiovisual percepts. Studies investigating the influence of selective adaptation on perception of auditory speech have demonstrated how adaptation to speech features can modulate perception of phonetic information in the auditory domain. These effects seem to occur even if the speech segments differ between the adaptor and the test stimuli. Returning to an example from the introduction, adaptation to auditory-/da/ can shift perception of phonetic categories along an auditory /ba/-to-/pa/ continuum towards /ba/ (e.g., Eimas & Corbit, 1973). It may be the case that adaptation effects in the visual domain are also influenced by feature-level information. For example, adaptation to visible speech articulation rate, previously found to modulate perception of auditory phonetic information (Green & Miller, 1985), could modulate visual speech perception even if the initial segments differ between the adaptor and test stimuli.

Alternatively, the information adapted in the visual mode may be specific to low-level sensory information (e.g., luminance, shape, and motion) as opposed to speech-specific phonetic information. Previous evidence demonstrating how adaptation to non-speech sounds (e.g., white noise) can modulate perception of auditory speech (e.g., Kat & Samuel, 1984) suggest that the adapted information need not be speech in nature to modulate subsequent perception of phonetic information. It is yet unknown whether adaptation to similar non-speech information can modulate perception of visual speech. Future research should investigate these possibilities.

The question of what information is modulated at the level of selective adaptation is made more complicated by reports of changes in auditory speech perception following adaptation to illusory phonetic information resolved from lexical context (e.g., Samuel, 1997, 2001; Samuel & Lieblich, 2014). For example, adaptation to auditory word-utterances containing a critical consonant (i.e. /b/ or /d/) can shift perceived phonetic boundaries along an auditory /bI/-to-/dI/ continuum. What is particularly interesting however is that the same perceptual changes are observed even when the critical consonant is replaced with noise. Because they are presented in the context of a word, these stimuli are typically perceived as still containing the missing consonant when in fact there is no sensory information for the consonant. The fact that these stimuli can induce selective adaptation suggests that there are cases for which common sensory-specific information is not required between adaptors and targets.

Samuel and Lieblich (2014) hypothesized that the reason McGurk-type adaptors fail to change auditory speech perception, yet illusory phonetic percepts derived from lexical context can, is due (in part) to competing phonetic information between the auditory and visual signals. However, stimuli producing visually influenced phonetic percepts without competing phonetic components (e.g., Green & Norrix, 2001) failed to produce changes to auditory speech perception equivalent to those produced by lexically-induced illusory phonetic percepts (Samuel & Lieblich, 2014). Based on this evidence, Samuel and Lieblich (2014) propose an admittedly speculative explanation. They propose that the influence visual context can have over auditory speech perception serves as a perceptual object, but not a linguistic object, while lexical context can serve as both. They seem to suggest that lexical context can provide more information than visual context during adaptation. Future research should explore the role of visible (lipread) lexical information on visual speech processing to determine if adaptation to lexical information can modulate visual speech perception similar to how lexical information can modulate auditory speech perception.

Conclusion

Within the current investigation, we observe that adaptation to salient visual speech information can modulate perception of audiovisual speech, based on sensory-specific influences. The results of the current investigation broaden understanding of the influence selective adaptation has on speech processing. The results demonstrate how speech adaptors that share sensory-specific phonetic information with audiovisual test stimuli can modulate the degree to which speech information provided by one sensory modality influences perception of speech in another sensory modality. The sensory-specific nature of the adaptation effects suggests that auditory and visual speech information are not completely integrated at the level of selective adaptation.

There are still many unanswered questions regarding what information is modulated at the level of selective adaptation. Current explanations for relevant information forms do not adequately account for the various findings within the literature. This is especially true in the face of the broader literature regarding multisensory speech processing and adaptation effects resulting from illusory phonemic information resolved from lexical context. A challenge for future research will be to account for these findings under a unifying explanation.

References

Arnold, P., & Hill, F. (2001). Bisensory augmentation: a speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92*(2), 339–355.

Baart, M., & Vroomen, J. (2010). Do you see what you are hearing? Cross-modal effects of speech sounds on lipreading. *Neuroscience Letters, 471*, 100–103.

Bernstein, L. E., Auer, E. T. J., & Moore, J. K. (2004). Audiovisual speech binding: Convergence or association. In G. A. Calvert (Ed.), *The handbook of multisensory processes* (pp. 203–223). Cambridge, MA: MIT Press.

Callan, D. E., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. *Cognitive Brain Research, 10,* 349–353.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., … David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*, 593–596.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10*(11), 649–657.

Campbell, C. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society B, 363*, 1001–1010.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. *Clinical Neurophysiology, 113*, 495–506.

Dias, J. W., Cook, T. C., & Rosenblum, L. D. (2016). The McGurk effect and the primacy of multisensory perception. In A. G. Shapiro & D. Todorovic (Eds.), *Oxford Compendium of Visual Illusions*. Oxford University Press (In press).

Eimas, P. D., Cooper, W. E., & Corbit, J. D. (1973). Some properties of linguistic feature detectors. *Perception & Psychophysics, 13*(2), 247–252.

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

Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders, 40*(4), 481–492.

Finney, D. J. (1971). *Probit analysis*. Cambridge, MA: Cambridge University Press.

Fowler, C. A. (2004). Speech as a supramodal or amodal phenomenon. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processing* (pp. 189–202). Cambridge, MA: MIT Press.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17*(3), 816–828.

Ganong, W. F. (1978). The selective adaptation effects of burst-cued stops. *Perception & Psychophysics, 24*(1), 71–83.

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature, 426*, 502–504.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279.

Green, K. P. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In R. Campbell, & B. Dodd (Eds.), *Hearing by Eye II: advances in the psychology of speechreading and audiovisual speech* (pp. 3–25). Hove, UK: Psychology Press.

Green, K. P., & Miller, J. L. (1985). On the role of visual rate information in phonetic perception. *Perception & Psychophysics, 38*(3), 269–276.

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.

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences, 106*(4), 1245–1248.

Jones, B. C., Feinberg, D. R., Bestelmeyer, P. E. G., DeBruine, L. M., & Little, A. C. (2010). Adaptation to different mouth shapes influences visual perception of ambiguous lip speech. *Psychonomic Bulletin & Review, 17*(4), 522–528.

Kat, D., & Samuel, A. G. (1984). More adaptation of speech by nonspeech. *Journal of Experimental Psychology: Human Perception and Performance, 10*(4), 512–525.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24*(3), 253–257.

Mallick, D. B., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type.

Psychonomic Bulletin & Review, 1–9, http://dx.doi.org/10.3758/s13423-015-0817-4.

Massaro, D. (1987). *Speech perception by ear and eye: a paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

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

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics, 72*(6), 1614–1625.

Musacchia, G., Sams, M., Nicol, T., & Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Experimental Brain Research, 168*(1–2), 1–10.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*(4), 2382–2393.

Pekkola, J., Ojanen, V., Autti, T., Jaaskelainen, I. P., Mottonen, R., Tarkiainen, A., & Sams, M. (2005). Primary auditory cortex activation by visual speech: An fMRI study at 3T. *Auditory and Vestibular Systems, 16*(2), 125–128.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd, & R. Campbell (Eds.), *Hearing by eye: the psychology of lip-reading* (pp. 97–113). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Remez, R. E. (2005). Perceptual organization of speech. In D. B. Pisoni, & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 28-50). Oxford: Blackwell.

Remez, R. E., Fellowes, J. M., Pisoni, D. B., Goh, W. D., & Rubin, P. E. (1998). Multimodal perceptual organization of speech: Evidence from tone analogs of spoken utterances. *Speech Communication, 26*, 65–73.

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

Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. Pisoni, & R. Remez (Eds.), *Handbook of Speech Perception* (pp. 51–78). Malden: Blackwell.

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science, 17*(6), 405–409.

Rosenblum, L. D., & Saldaña, H. M. (1992). Discrimination tests of visually influenced syllables. *Perception and Psychophysics, 52*(4), 461–473.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17*, 1147–1153.

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

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S., & Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127*, 141–145.

Sams, M., Mottonen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research, 23*, 429–435.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32*, 97–127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*(4), 348–351.

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), 1479–1490.

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

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In I. Barbara, & C. Ruth (Eds.), *Hearing by eye: the psychology of lip-reading* (pp. 3–51). Hillsdale, NJ: Erlbaum.

Thomas, S. M., & Jordan, T. R. (2002). Determining the influence of Gaussian blurring on inversion effects with talking faces. *Attention, Perception, & Psychophysics, 64*(6), 932–944.

Vatikiotis-Bateson, E., Eigsti, I., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics, 60*(6), 926–940.

Vroomen, J., & Baart, M. (2012). Phonetic recalibration in audiovisual speech. In M. M. Murray, & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes* (pp. 363–379). Boca Raton, FL: CRC Press.

Table 1.1

Hypothesized changes in categorization of audiovisual test-stimuli following adaptation

Adaptor	Adaptation to Cross-Sensory Phonetic Information	Adaptation to Sensory-Specific Phonetic Information
Auditory-/va/	More continuum items identified as /ba/	No change
Visual-/va/	More continuum items identified as /ba/	More continuum items identified as /ba/
Audiovisual-/va/	More continuum items identified as /ba/	More continuum items identified as /ba/
Auditory-/ba/	More continuum items identified as /va/	No change
Visual-/ba/	More continuum items identified as /va/	No change

Notes. "Adaptation to Cross-Sensory Phonetic Information" assumes adaptation affects perception of integrated audio-visual speech by affecting processing of crossmodal phonetic information. "Adaptation to Sensory-Specific Phonetic Information" assumes adaptation affects perception of integrated audio-visual speech by affecting processing of sensory specific information shared between the adaptor and audio-visual test-stimuli.

Table 1.2

Phonemic boundary shifts following adaptation

Adaptor	Baseline	Adapted	Shift	*SE*	*t*	*n*	*p*	*r*
A-/va/	4.869	4.551	-0.318	1.594	-0.199	10	.423	.063
V-/va/	3.486	2.100	-1.386	0.448	-3.095	10	.007	.699
AV-/va/	3.887	1.900	-1.987	0.661	-3.007	10	.008	.689
A-/ba/	4.207	5.134	0.927	0.678	1.367	10	.103	.397
V-/ba/	5.545	5.718	0.173	0.827	0.210	10	.419	.066

Note: Baseline and adapted values represented the cumulative normal ogives for the hypothetical test-stimulus corresponding to the 50% point, representing the average phonemic boundary of the test-continuum before and following adaptation. t-values represent 1-tailed paired-samples tests. A negative shift denotes a phonemic boundary shift towards the /va/-end of the continuum, indicating more "ba" responses following adaptation.

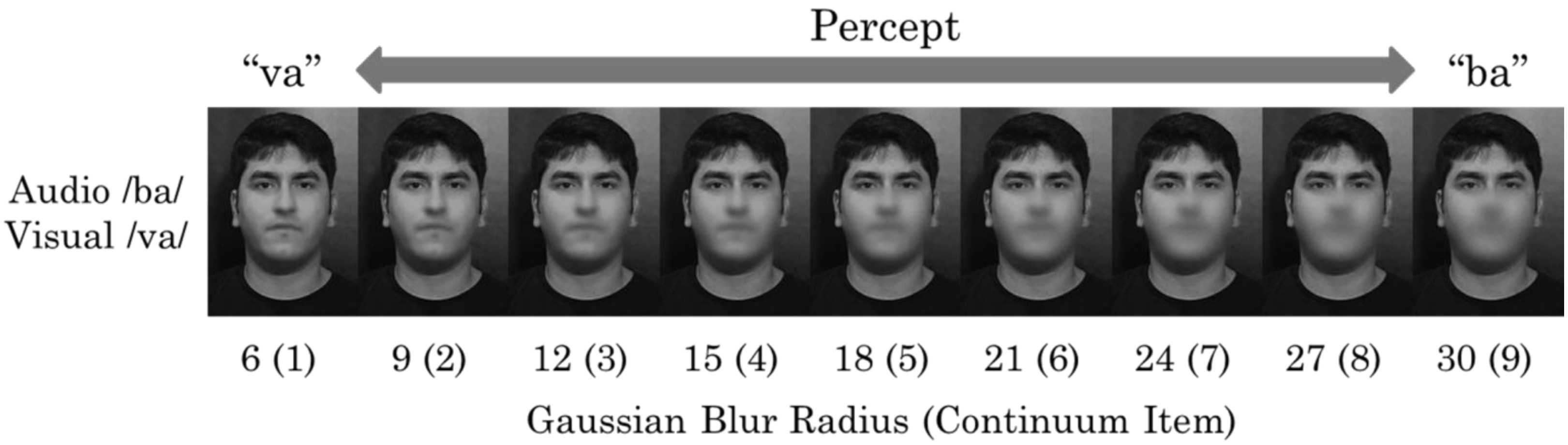


Figure 1.1. The nine-item audiovisual test continuum of integrated phonetic percepts. As Gaussian Blur becomes stronger, the salience of the visual information becomes less. However, for all items in the continuum, the auditory component remains the same (/ba/). The strength of the McGurk illusion becomes weaker as visibility of the mouth decreases. As a result, greater reliance is put on the auditory component of the audiovisual stimulus, decreasing perception of the illusory "va" percept.

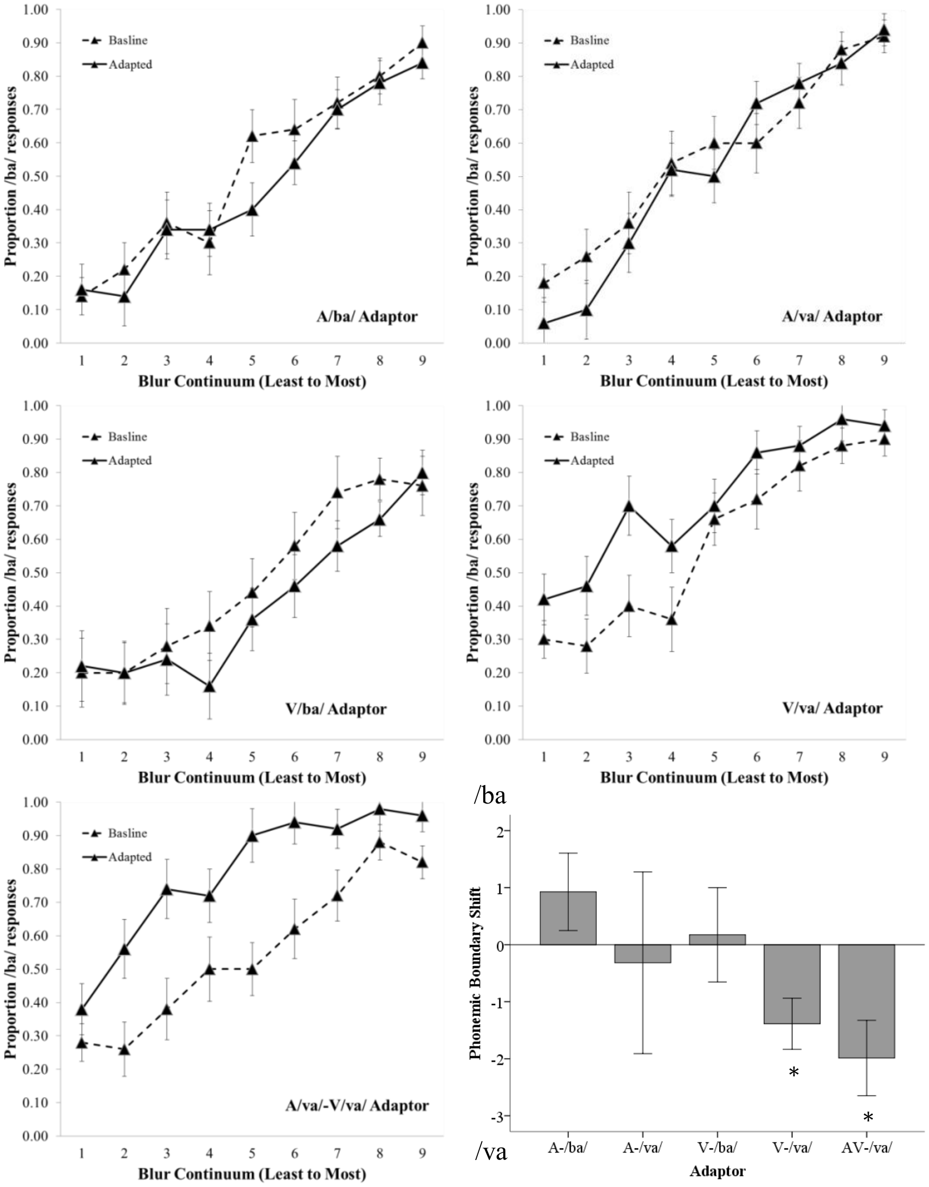


Figure 1.2. The proportion of /ba/ responses for each of the nine test-continuum items prior to (Baseline) and post adaptation (Adapted) for each of the four adaptors. The bottom left panel illustrates the phonemic boundary shift for each adaptor. Positive values denote shifts towards the /ba/-end of the continuum. Negative values denote shifts towards the /va/-end of the continuum. Error bars represented the standard error of the mean.

Chapter 2

Selective adaptation of crossmodal speech information:

A case of small but consistent effects

Speech is multimodal (for a review, see Rosenblum, 2008). Visual speech information can improve identification of auditory speech heard in difficult listening conditions (e.g., Erber, 1975; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954) or when auditory speech is reduced in acoustic quality (e.g., Pilling & Thomas, 2011; Remez, Fellowes, Pisoni, Goh, & Rubin, 1998). Visual speech information can also improve comprehension of accented speech (e.g., Navarra & Soto-Faraco, 2007; Sueyoshi & Hardison, 2005) or speech that conveys complicated content (e.g., Arnold & Hill, 2001; Reisberg, McLean, & Goldfield, 1987). Perceivers will even non-consciously imitate the subtle articulatory characteristics of speakers they hear (e.g., Goldinger, 1998; Pardo, 2006) and lipread (Miller, Sanchez, & Rosenblum, 2010), demonstrating how audible and visible speech can modulate speech production. Work in our own lab has also demonstrated how familiarity with speech spoken by a talker in one sensory modality can improve later comprehension of that same talker's speech in a different sensory modality (Rosenblum, Miller, & Sanchez, 2007; Sanchez, Dias, & Rosenblum, 2013). The literature clearly demonstrates how auditory and visual sources of information can influence speech processing and how information available in one stream can influence processing of information in another stream.

Perhaps the most striking demonstrations of this crossmodal influence in speech processing come from examples where conflicting information provided across sensory modalities can *change* perception of an acoustic phonetic signal to some other percept. For example, the McGurk effect (McGurk & MacDonald, 1976) demonstrates how dubbing an auditory speech signal (e.g., auditory-/ba/) with an inconsistent visual speech signal (e.g., lipread-/ga/), can produce an audiovisual stimulus perceived as an integration of the conflicting sensory components (e.g., heard as “da”). Similar crossmodal influences on auditory speech perception have been demonstrated using tactile (e.g., Derrick & Gick, 2013; Fowler & Dekle, 1991; Gick & Derrick, 2009), kinematic (e.g., Rosenblum & Saldaña, 1996), and kinesthetic (e.g., Ito, Tiede, & Ostry, 2009; Sams, Mottonen, & Sihvonen, 2005) sources of information.

The breadth of the crossmodal literature suggests that information important for speech may take a form accessible across sensory modalities (e.g., Fowler, 2004; Rosenblum, 2005; Rosenblum, Dias, & Dorsi, 2016). However, some studies of crossmodal influence in the domain of *selective speech adaptation* seem to suggest that speech information is processed with some difference between the senses (Dias, Cook, & Rosenblum, 2016; Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014). However, as we will discuss in the following section, there is some discrepancy within the selective speech adaptation literature regarding crossmodal influence that we feel warrants further investigation.

Investigations of crossmodal influence in selective speech adaptation

Selective speech adaptation describes a perceptual aftereffect following repeated exposure (adaptation) to clear phonetic information. For example, in their seminal study, Eimas and Corbit (1973) demonstrated how adaptation to a repeated auditory utterance of

/ba/ (i.e., hearing /ba/ repeated 75 times in a row) can cause perceivers to then identify fewer test-stimuli along a 14-step auditory /ba/-to-/pa/ continuum as sounding like /ba/.

The classic literature on selective speech adaptation clearly demonstrates how adaptation to auditory speech information can change auditory speech perception (e.g., Eimas & Corbit, 1973; Ganong, 1978; Pisoni & Luce, 1987). Recent research also demonstrates how adaptation to visual speech information can change visual speech perception. Baart and Vroomen (2010) found that adaptation to an audio-video recording of a talker dynamically articulating /omso/ or /onso/ can change how perceivers categorize silent videos along an /omso/-to-/onso/ continuum of dynamic articulating faces. Perceivers identified more videos as "onso" following adaptation to audiovisual /omso/ and more videos as "omso" following adaptation to audiovisual /onso/. Jones, Feinberg, Bestelmeyer, DeBruine, and Little (2010) also demonstrated how adaptation to still images of faces clearly articulating /m/ or /u/ can change how perceivers categorize faces on an /m/-to-/u/ continuum of still-face images. Following adaptation to /m/, perceivers identified more continuum items as /u/, and vice versa.

Returning to multisensory speech, evidence for the crossmodal influences discussed previously has motivated researchers to investigate whether adaptation to speech in one sensory modality can change perception of speech in another modality (Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014). Early on, Cooper and colleagues did observe some evidence for adaptation between the auditory and *motor* systems (Cooper, Billings, & Cole, 1976; Cooper, Blumstein, & Nigro, 1975; Cooper & Lauritsen, 1974). While the motor system is not typically considered a perceptual modality, as such, it does involve sensations (i.e. kinesthesis and touch) that can guide speech production. Cooper and Lauritsen (1974) observed that the voice-onset times of perceivers' productions of /pi/ were shortened (becoming more /bi/-like) following selective adaptation to an auditory utterance of /pi/. In subsequent investigations, Cooper and colleagues found analogues influences occurring in the opposite direction (Cooper et al., 1976; Cooper et al., 1975). For example, perceivers identified fewer test-items along a /bæ/-/dæ/-/gæ/ auditory continuum as /bæ/ after repeatedly producing /bæ/ utterances (controlling for possible influences of hearing one's own voice; see Cooper et al., 1975, for details). These investigations demonstrate a bidirectional adaptation effect between auditory speech perception and speech production: Heard speech can influence speech production, and speech production (and the kinesthetic and tactile sensations associated with it) can induce changes in auditory speech perception.

Following Cooper and colleagues' studies, Roberts and Summerfield (1981) tested for auditory-*visual* crossmodal selective speech adaptation. They addressed the question of whether selective speech adaptation is based on phonetic or auditory features/primitives. They hypothesized that if phonetic information, which is modality neutral (e.g., can be heard and lipread), can adapt across sensory modalities, then adaptation to visual (lipread) speech should change how auditory speech is heard. However, following adaptation to a repeated video of a talker silently articulating /bɛ/ or /dɛ/, perceivers failed to exhibit a significant change in their identification of speech along an auditory /bɛ/-to-/dɛ/ continuum. Roberts and Summerfield (1981) concluded that modality-neutral phonetic information is insufficient to induce crossmodal adaptation and suggest that selective speech adaptation depends on the sense-specific (auditory) spectrotemporal relationship between adaptors and test-stimuli. Roberts and Summerfield (1981) do not reconcile their results with the evidence suggesting repeated speech production can change auditory speech identification (Cooper, 1974; Cooper et al., 1976; Cooper et al., 1975). However, it is interesting to note that despite being a non-significant change, Roberts & Summerfield's (1981) data did show that more auditory test-stimuli were identified as /dɛ/ following adaptation to visual-/bɛ/ and more were identified as /bɛ/ following adaptation to visual-/dɛ/. This change was quite small and much less than the change observed following auditory speech adaptation. This point will be addressed later.

The conclusion of Roberts and Summerfield (1981) that selective adaptation is based on sense-specific (e.g., auditory) information is consistent with experiments testing *bimodal* audio-visual speech adaptation. For example, in the same paper, Roberts and Summerfield (1981) found that changes in auditory speech identification following bimodal audiovisual adaptation were no greater than changes following auditory adaptation, suggesting there was no influence of redundant crossmodal information. Further, they reported another experiment in which more auditory test-stimuli were identified as /dɛ/ following adaptation to an auditory-/bɛ/-visual-/dɛ/ bimodal stimulus, an audiovisual combination that is typically perceived as /dɛ/ (e.g., McGurk & MacDonald, 1976). What they observed was a change in auditory speech identification that reflected adaptation to the auditory component of the audio-visual adaptor instead of adaptation to the audio-visual /dɛ/ phonetic perception (which would have produced more /bɛ/

responses). Saldaña and Rosenblum (1994) later observed similar effects when employing an auditory-/ba/-visual-/va/ adaptor, which produced a very compelling visually influenced “va” percept 99% of the time. Despite the more compelling stimulus adaptor, perceivers still exhibited a change in auditory test-stimulus identification consistent with adaptation to the auditory-/ba/ component of the audio-visual adaptor, not the audio-visual “va” percept.

Recently, Dias et al. (2016) used an incongruent audio-/ba/-visual-/va/ stimulus, to construct an audiovisual *test*-continuum. By gradually obstructing visibility of the articulating mouth across 9 steps, the test-continuum ranged from a clear visually-influenced “va” percept to a clear auditory-dependent “ba” percept. Changes in audiovisual speech categorization only followed adaptation to visual-/va/ and audiovisual-/va/ (not auditory-/va/ or visual-/ba/). Further, there was no difference in adaptation between adaptors, suggesting that redundant phonetic information provided across modalities (auditory-/va/ in the audiovisual-/va/ adaptor) does not increase adaptation. The results again suggest that there must be shared sensory information between adaptors and test-stimuli (visual in this case) to observe adaptation, consistent with Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994). To reconcile these findings with the large literature demonstrating crossmodal influences in other speech domains, Saldaña and Rosenblum (1994) and Dias et al. (2016) suggest that information across the auditory and visual modalities is not (completely) integrated at the level of selective adaptation. As such, selective adaptation effects may depend primarily on sensory mechanisms at the earliest levels of perceptual processing (Dias et al., 2016).

Roberts and Summerfield (1981), Saldaña and Rosenblum (1994), and Dias et al. (2016) all suggest that selective speech adaptation depends (primarily) on shared sensory information between adaptors and test-stimuli. However, other work in selective speech adaptation suggests that the sensory information shared between adaptors and test-stimuli can be featural in nature. For example, Eimas and Corbit (1973) in their seminal investigation observed how changes in the identification of stimuli along an auditory /ba/-to-/pa/ continuum can follow adaptation to auditory-/da/ or auditory-/ta/. Fewer test stimuli were identified as /ba/ following adaptation to /da/, and fewer test stimuli were identified as /pa/ following adaptation to /ta/. Eimas and Corbit (1973) suggested that because similar voice-onset time characteristic are shared between /b/ and /d/, and between /p/ and /t/, that the resulting adaptation to voice-onset features indicates that selective adaptation is sensitive to sub-phonetic, featural speech information. Still other evidence suggests that sensory information shared between adaptors and test-stimuli need not be speech in nature (e.g., white noise), so long as there is spectrotemporal overlap between them (e.g., Diehl, Kluender, & Parker, 1985; Kat & Samuel, 1984; Samuel & Newport, 1979). However, there are instances of selective adaptation for which changes in speech perception can follow adaptors that do not share spectrotemporal information with test-stimuli.

Lexical influences in selective speech adaptation

Samuel and his colleagues (Samuel, 1997, 2001; Samuel & Lieblich, 2014) have found that when the critical consonant in an auditory word is replaced with noise, the word-utterance is typically perceived as intact, with the consonant still present. This perceptual illusion is known as the *phonemic restoration effect* (e.g., Samuel, 1981). Interestingly, if these restored words are used as adaptors, auditory speech identification can change as if the consonant was actually present in the adaptor (e.g., Samuel, 1997, 2001). For example, more stimuli along an auditory /ba/-to-/da/ continuum are identified as /ba/ following adaptation to an auditory utterance of “armadillo” that has had the critical /d/ segment removed and replaced with noise (e.g., Samuel, 1997). These studies seem to suggest that an adaptor need not share featural or sensory information with test-stimuli for selective speech adaptation to occur.

To reconcile these results with the bimodal findings of Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994), Samuel and Lieblich (2014) speculate on a kind of hierarchy of information that can induce selective speech adaptation. They suggest that adaptation is first influenced by clear bottom-up, sense-specific information – as occurs when auditory adaptors influence identification of stimuli along an associated auditory continuum (e.g., Eimas & Corbit, 1973). Samuel and Lieblich (2014) go on to suggest that in the absence of sense-specific information, top-down information (i.e., visual or lexical/linguistic context) should drive adaptation. The top-down influence of lexica/linguistic context can explain why restoration-based adaptors can change auditory speech identification. Further, Samuel and Lieblich (2014) suggest that the reason why the bimodal (McGurk-type) audio-visual adaptors employed by Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994) produced adaptation effects relative to their auditory components is because the auditory components provided sense-specific information salient to the auditory test-stimulus continua. Because sense-specific information relevant to the test continua was available to drive adaptation, the integrated audio-visual phonetic percepts did not.

To reconcile their findings with the lack of *cross*modal adaptation observed by Roberts and Summerfield (1981), and their own lack of finding changes in speech identification following adaptation to visually influenced auditory percepts derived from non-competing audio-visual components (using a technique adapted from Green & Norrix, 2001), Samuel and Lieblich (2014) propose an important distinction between visual speech information and lexical information. They suggest that lexically derived phonetic information can induce selective speech adaptation (in the absence of bottom-up sense-specific information) because it provides both perceptual (phonetic) and linguistic (lexical) information. They go on to suggest that the influence of visual speech on auditory speech perception is perceptual only. They propose that, "One way to think about this distinction is to note that from a linguistic perspective, [auditory] phonetic segments are naturally associated with the lexical representations that contain them; in contrast, the visual pattern that is associated with a word does not have the part–whole relationship of segments and words." (Samuel & Lieblich, 2014, p.1488). They offer as support for this distinction other evidence from the speech literature that seems to suggest auditory linguistic information can trump visually influenced speech percepts. For example, presentation of an audio-/beef/-visual-/deef/ stimulus will induce semantic priming effects for words associated with the word "beef" (e.g., pork), even though perceivers report hearing the audiovisual stimulus as "deef" (Ostrand, Blumstein, Ferreira, & Morgan, 2016). Samuel and Lieblich (2014) suggest that noise-replaced word adaptors can change auditory speech identification because they provide both perceptual and linguistic information, as opposed to visual speech, which they argue can only provide perceptual information.

Interestingly, Samuel and Lieblich (2014) do not reconcile their description of visual speech as non-linguistic with evidence suggesting visual speech *can* behave in a linguistic manner. For example, Kim, Davis, and Krins (2004) observed repetition priming effects when lipread primes preceded auditory utterances of the same word, but not when lipread primes preceded auditory utterances of the same *non*-words. There is also a growing body of evidence suggesting visual speech information can activate cortical areas associated with lexical processing (for a review, see Bernstein & Liebenthal, 2014). Further, like Roberts and Summerfield (1981), Samuel and Lieblich (2014) do not reconcile their results with the aforementioned evidence suggesting repeated speech production can change auditory identification (Cooper, 1974; Cooper et al., 1976; Cooper et al., 1975). Based on their theoretical account distinguishing between perceptual and linguistic influences in selective speech adaptation, it is unclear why repeated speech production should change auditory speech perception (Cooper, 1974; Cooper et al., 1976; Cooper et al., 1975) while visual-alone speech adaptation does not (Roberts & Summerfield, 1981).

Summary of the problem

Evidence from studies investigating bimodal audio-visual speech adaptation suggest that sensory information shared between adaptors and test-stimuli will determine selective speech adaptation over phonetic percepts (e.g., Roberts & Summerfield, 1981;

Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014). Further, adaptation to visual speech information does not significantly change auditory speech identification (Roberts & Summerfield, 1981). However, adaptation to phonetic percepts determined by lexical context (i.e., phonemic restoration effect) *can* change perception of auditory speech (Samuel, 1997, 2001; Samuel & Lieblich, 2014).

Arguably, the theoretical account offered by Samuel and Lieblich (2014) for why lexically-influenced phonetic percepts can affect selective speech adaptation, while visually-influenced phonetic perceptions and visual speech alone cannot, is unsatisfactory. Their description of lexically-influenced phonetic percepts as both perceptual and linguistic while visually-influenced phonetic percepts and visual speech are perceptual only is inconsistent with evidence in the speech literature demonstrating how visual (lipread) speech can exhibit lexical qualities (Bernstein & Liebenthal, 2014; e.g., Kim et al., 2004; but see Ostrand et al., 2016). Further, the theoretical account offered by Samuel and Lieblich (2014) does not provide an explanation for why repeated speech production (and the kinesthetic and tactile sensations associated with it) can change auditory speech identification (Cooper et al., 1976; Cooper et al., 1975).

Given the literature in other speech domains for crossmodal influence (for reviews, see Rosenblum, 2008; Rosenblum et al., 2016), it is odd not to find audio-visual crossmodal adaptation (Roberts & Summerfield, 1981). It is especially odd considering Cooper and colleagues demonstrated adaptation between the auditory and motor systems, suggesting a crossmodal influence between audition and kinesthesis/touch (Cooper, 1974; Cooper et al., 1976; Cooper et al., 1975). It is odd also considering evidence for auditory-visual crossmodal adaptation outside of the speech domain. For example, following adaptation to visual horizontal motion, perceivers identify stationary auditory sounds as moving in the opposite direction (Ehrenstein & Reinhardt-Rutland, 1996). Similarly, adaptation to visual motion in depth transfers to an auditory motion aftereffect (perceived changes in sound intensity), though adaptation to auditory motion in depth does not transfer to a visual motion aftereffect (Kitagawa & Ichihara, 2002). More recently, bidirectional aftereffects have been demonstrated for stimulus presentation rate between auditory and visual stimuli (Levitan, Ban, Stiles, & Shimojo, 2015).

Current investigation

Roberts and Summerfield (1981) provide the *only* published investigation of crossmodal selective speech adaptation between the auditory and visual modalities. As has been previously discussed, most auditory-visual investigations employ bimodal audiovisual adaptors (e.g., Baart & Vroomen, 2010; Dias et al., 2016; Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014). Given that there is only this one attempt published in the literature for crossmodal adaptation between auditory and visual speech, further investigation is certainly warranted. This seems especially true when one considers (as discussed above) that the non-significant observations made by Roberts and Summerfield (1981) may suggest reliable, though small, crossmodal adaptation. As such, the purpose of the current investigation is to further explore whether adaptation to speech information in one modality, either auditory or visual, can change identification of phonetic information in the other modality.

Experiment 1

The goal of Experiment 1 is to attempt to determine whether adaptation to visual utterances of /ba/ or /va/ can change identification of speech in an auditory /ba/-to-/va/ test-continuum (e.g., Roberts & Summerfield, 1981). Further, Experiment 1 also seeks to determine whether adaptation to auditory utterances of /ba/ or /va/ can change identification of speech in a visual /ba/-to-/va/ test-continuum. Along with these crossmodal adaptation paradigms, we also replicate within-modality adaptation, measuring changes in auditory speech identification following auditory speech adaptation and changes in visual speech identification following visual speech adaptation. Roberts and Summerfield (1981) failed to find significant auditory-visual crossmodal influences employing /b/ and /d/ adaptors and test-stimuli. This may be due in part to the fact that /b/ and /d/ differ only in place of articulation (bilabial vs. alveolar, respectively). However, /b/ and /v/ differ in both place (bilabial vs. labiodental) and manner (stop vs. fricative) of articulation, making these phonemes more discriminable. In fact, the differences in place of articulation between /b/ and /v/ have been found to make these phonemes particularly distinct *visually* (e.g., Rosenblum & Saldaña, 1992, 1996; Saldaña & Rosenblum, 1994).

If selective speech adaptation can be influenced by information available across the auditory and visual modalities, then adaptation to visual speech should change auditory speech identification and vice versa. However, if selective speech adaptation depends on sense-specific spectrotemporal overlap between adaptors and test-stimuli, then adaptation to visual speech should not change auditory speech identification and vice versa.

Method

Participants. Eighty undergraduates (42 female, 38 male; M_{age} = 19.16 years, SE_{age} = .103) from the University of California, Riverside, participated in partial fulfillment of course credit. All participants were native English speaking with normal hearing and normal or corrected-to-normal vision. These participants were randomly and evenly distributed among eight different groups depending on the adaptor and test-continuum used.

Materials.

Test continua. Two male speakers (both 32-year-old native English speaking; Speaker S1 a California native, Speaker S2 an Indiana native) were digitally audio-video recorded uttering /ba/ and /va/. Recordings were made at 30 frames-per-second at a size of 640x480 pixels and 16-bit sound. Audio-video editing was executed using Final Cut Pro for Mac OSX and Praat (Boersma & Weenink, 2013).

Auditory test continua. Adapting a method employed by Saldaña and Rosenblum (1993), an eleven-item auditory /va/-to-/ba/ test continuum was made for each speaker by first digitally extracting the auditory components of their /va/ utterances. This /va/ was then made into a /va/-to-/ba/ continuum by systematically deleting pieces of the /va/ fricative and replacing it with silence. This was accomplished by first finding the point along the initial frication of a /va/ utterance judged by the experimenters to be most ambiguous between /va/ and /ba/ (half of the time perceived as “va” and half of the time perceived as “ba”). This point served as the midpoint of a speaker’s continuum. Five equal steps before (longer fricative) and after (shorter fricative) the midpoint were determined to ensure that those steps before the midpoint ended with an unambiguous /va/ sound and those steps after the midpoint ended in an unambiguous /ba/ sound. Once these steps were determined, the initial /va/ utterance was copied eleven times. The initial fricative of each copy was replaced with silence up to one of the previously determined steps. For speaker S1, the initial fricative was reduced in 3ms increments per step. For speaker S2, the initial fricative was reduced in 7ms increments per step. The finished stimuli were each 1.604s in length. The resulting continuum for each speaker was pilot tested and found to be unambiguously identified as /va/ at the end of the continuum with the longest fricative and unambiguously identified as /ba/ at the end of the continuum with the shortest fricative, with the middle tokens identified half of the time as /va/ and half of the time as /ba/.

Visual Test Continua. Adapting a method employed by Baart and Vroomen (2010), an eleven-item visual /va/-to-/ba/ test continuum was made for each speaker by first digitally extracting the visual components of the speaker's /va/ and /ba/ utterances. Each speaker's /va/ utterance was then digitally superimposed over their /ba/ utterance. The two video images lined-up well when superimposed and did not require any adjustment of video position (see *Figure 1*). However, fine temporal adjustments were made to ensure that the visible articulatory gestures for both the /ba/ and /va/ utterances began at the same time. After adjusting to line up the initiation of articulation, clips for the /ba/ and /va/ videos finished within a single frame of one another and required no adjustment of video length.

Eleven copies of the stimulus were then created. Each of the eleven copies was then modified to decrease the opacity of the superimposed /va/ utterance in 10% increments. The finished stimuli were each 1.604s in length, corresponding with the auditory test continua. Figure 1 provides examples of still frames taken at the same point in each of the 11 stimulus videos. As seen, with the decreasing opaqueness of the /va/ utterance, the /ba/ utterance becomes more visible. The continuum for each speaker was pilot tested and found to be unambiguously identified as /va/ at the end of the continuum where opacity of the superimposed /va/ utterance was 100% and unambiguously identified as /ba/ at the end of the continuum where the opacity of the superimposed /va/ utterance was 0%, with the middle tokens (50%) identified half of the time as /va/ and half of the time as /ba/.

Adaptors. Unimodal auditory and visual adaptors were extracted from the auditory and visual continua: The /va/-endpoints of the auditory test continua served as auditory-/va/ adaptors, while the /ba/-endpoints of the auditory test continua served as auditory-/ba/ adaptors. Similarly, the /va/-endpoints of the visual test continua served as visual-/va/ adaptors, and the /ba/-endpoints of the visual test continua served as visual-/ba/ adaptors. Silent non-articulating frames before and after the CV utterances were systematically removed, reducing the length of the adaptor stimuli to 734ms. We have successfully used this technique before to more efficiently adapt participants, presenting more adaptor repetitions across a shorter amount of time (e.g., Dias et al., 2016).

Procedure. The different adaptation conditions were tested as a between-groups factor (e.g., Dias et al., 2016; Samuel & Lieblich, 2014). The 80 participants were equally divided into eight experimental groups, depending on the continuum they were tested on and the adaptor they were exposed to. The modality-consistent adaptation groups included: 1) auditory test continuum and auditory-/ba/ adaptor; 2) auditory test continuum and auditory-/va/ adaptor; 3) visual test continuum and visual-/ba/ adaptor; and 4) visual test continuum and visual-/va/ adaptor. The crossmodal adaptation groups included: 1) auditory test-continuum and visual-/ba/ adaptor; 2) auditory test continuum and visual-/va/ adaptor; 3) visual test continuum and auditory-/ba/ adaptor; and 4) visual test continuum and auditory-/va/ adaptor. Within these eight groups, half (5) of the participants listened to tokens derived from Speaker S1 and the other half listed to tokens derived from Speaker S2.

Participants initially identified members of the eleven-item test continuum specific to their condition group. Identification trials initiated with a crosshair presented on a computer screen for 500ms followed by presentation of one test stimulus. Auditory stimuli were presented while the crosshair remained on the screen. Visual stimuli were presented by replacing the crosshair after the initial 500ms interval. Following stimulus presentation, participants indicated whether they heard/lipread “ba” or “va” by pressing a keyboard key labeled with the corresponding response. Participants identified each of the eleven test stimuli twelve times in random order for 132 identifications. This baseline measure was collected to determine where participants perceived a phonemic category boundary along the corresponding test continuum before adaptation.

The adaptation phase of the experiment started with an initial adaptation consisting of 100 exposures to the adaptor. Following adaptation, participants identified three randomly selected test-continuum stimuli in the same manner as in the baseline task. Participants then completed 43 cycles, each consisting of 50 adaptor exposures followed by identification of three of the test stimuli. Using twice as many exposures to the adaptor in the initial block is a technique commonly used to build up adaptation before making identifications of test-stimuli (e.g., Dias et al., 2016; Eimas & Corbit, 1973; Roberts & Summerfield, 1981). Across these adaptation cycles, participants identified each test continuum stimulus twelve times for 132 identifications.

Participants were instructed to remain silent throughout the procedure. The entire experiment took approximately one hour to complete for each participant. The experimental procedure was executed using PsyScope Software for Mac OSX (Cohen, MacWhinney, Flatt, & Provost, 1993).

Results & Discussion

Participant identifications were coded as the proportion of times each of the test continuum stimuli were identified as /ba/ prior to and after adaptation (*see Figure 2*). Then, as with previous studies (e.g., Dias et al., 2016; Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994), cumulative normal ogives were fit to each participant’s identification function across the eleven-item test continuum before and after adaptation, employing the method of probits (Finney, 1971). The number for the hypothetical test stimulus corresponding to the 50% point of each participant’s function provided a measure of the perceived phonemic category boundary between /ba/ and /va/. Each experimental group was evaluated to determine whether the phonemic category boundary significantly changed after adaptation, using one-tailed paired-samples t-tests.

As illustrated in *Table 2.1*, groups that were adapted to speech stimuli in a modality consistent with the test stimuli (auditory-auditory, visual-visual) all demonstrated a significant phonemic category boundary shift following adaptation. The results are consistent with studies demonstrating changes to auditory phonetic perception following auditory speech adaptation (e.g., Diehl, 1975; Eimas & Corbit, 1973; Ganong, 1978; Vroomen & Baart, 2009) and with studies demonstrating changes to visual speech perception following visual speech adaptation (e.g., Baart & Vroomen, 2010; Jones et al., 2010).

However, groups that were adapted to speech stimuli in the modality different from the test stimuli (auditory-visual, visual-auditory) all failed to demonstrate a significant phonemic category boundary shift following adaptation. Visual speech adaptation failed to produce significant changes in categorization of auditory speech, and auditory speech adaptation failed to produce significant changes in categorization of visual speech.

The results of Experiment 1 seem to be consistent with those of Roberts and Summerfield (1981), demonstrating non-significant changes in speech identification following crossmodal adaptation of speech information. The fact that we failed to demonstrate significant crossmodal adaptation does not seem to be a result of poor stimuli, considering within-modality adaptation was quite robust, consistent with previous research for within-modality adaptation in auditory speech perception (e.g., Eimas & Corbit, 1973; Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994) and in visual speech perception (e.g., Baart & Vroomen, 2010; Jones et al., 2010; Vroomen & Baart, 2012).

However, recall that an evaluation of the data reported by Roberts and Summerfield (1981) suggests that though their observed shifts in phonemic categorization failed to reach statistical significance, the shifts track in directions predicted by their respective crossmodal adaptors. Adaptation to visual /bɛ/ shifted the auditory phonemic category boundary toward /bɛ/ and adaptation to visual /dɛ/ shifted the auditory phonemic category boundary toward /dɛ/, though neither shift reached statistical significance. Based on this observation, we examined whether a similar pattern occurred in the current data.

In fact, all but one of the crossmodal conditions tested in Experiment 1 produced subtle phonemic boundary shifts in the direction predicted by crossmodal adaptation (see *Table 2.1*). However, these crossmodal effects are smaller than their within-modality counterparts (see *Table 2.2*).

The pattern of non-significant crossmodal adaptation observed by Roberts and Summerfield (1981) and in Experiment 1 may indicate reliable, though small, crossmodal adaptation across studies. To evaluate this possibility, we conducted a meta-analysis including the two visual-to-auditory crossmodal adaptation conditions tested by Roberts and Summerfield (1981) with the two visual-to-auditory and two auditory-to-visual crossmodal adaptation conditions tested in Experiment 1. Because Roberts and Summerfield (1981) reported the data for each of their participants in each of their experiments, we have all necessary information for conducting a meta-analysis.

Meta-Analysis 1

We conducted the meta-analysis using a procedure described by Cumming (2012) to determine whether the apparent consistency of crossmodal selective adaptation effects across the conditions tested by Roberts and Summerfield (1981) and the conditions tested in Experiment 1 are significant across studies. First, we coded the phonemic boundary shifts for each condition observed by Roberts and Summerfield (1981) and ourselves from Experiment 1 as reflecting either a shift in the predicted direction (positive) or a shift in the non-predicted direction (negative). Each conditions weight reflects the proportion of each condition's inversed squared standard error to the sum of the inversed squared standard errors across conditions (e.g., Cumming, 2012). This weighting technique ensures that study conditions with more participants and less variability (accounting for amount and quality of information) have more weight when calculating the mean and effect size across study conditions. For example, though the visual-/va/ adaptor produced the greatest phonemic boundary shift along the auditory-/va/-to-/ba/ continuum (from Experiment 1), this condition also produced the greatest variability. As a result, this condition was weighted least (3%) in the meta-analysis. The overall standard error of the meta-analysis is computed as the square root of the inverse of the summed inversed squared standard errors across conditions (e.g., Cumming, 2012). We used the weighted mean and standard error to compute a 1-sample *t*-test with a null value of 0 (no change in phonemic boundary shift).

Results & Discussion

Table 2.3 reports the results of Meta-Analysis 1. The results reveal that the mean phonemic boundary shift across studies ($M = 0.138$, $SE = 0.082$) is significantly greater than 0, $t(51) = 1.681$, $p = .049$, $r = .253$. Further, a test of heterogeneity revealed that the phonemic boundary shifts observed by Roberts and Summerfield (1981) and ourselves did not significantly differ across studies and conditions, $\chi^2(5) = 1.358$, $p = .929$.

The results of Meta-Analysis 1 suggest that when the crossmodal adaptation effects observed by Roberts and Summerfield (1981) are combined with the crossmodal adaptation effects we observed in Experiment 1, overall crossmodal adaptation between the auditory and visual modalities (across auditory-to-visual and visual-to-auditory) is significant. These results suggest that though crossmodal speech adaptation was small in these studies, it was *reliable*.

That any single attempt to induce crossmodal adaptation in speech failed to reach statistical significance may be related to issues of statistical power, based on the number of participants in any particular group. Keep in mind that Roberts and Summerfield (1981) tested groups of only 6 participants, and we tested groups of only 10 participants, consistent with other such studies conducted in our lab (e.g., Dias et al., 2016; Saldaña & Rosenblum, 1994). Employment of such a low number of participants is common in the selective speech adaptation literature and seems sufficient to induce *within*-modality adaptation (e.g., Bertelson, Vroomen, & de Gelder, 2003; Cooper et al., 1976; Eimas & Corbit, 1973; Samuel & Kat, 1998). However, more recent studies have employed larger groups of participants (e.g., 18-32) when investigating what are presumed to be more subtle influences in adaptation, including those induced by lexical context (e.g., Samuel, 1997; Samuel & Lieblich, 2014; see also Vroomen & Baart, 2009; Vroomen, Van Linden, De Gelder, & Bertelson, 2007). It could be then that the differences in results (significant vs. non-significant changes in speech identification) observed between investigations of crossmodal adaptation (Experiment 1 and Roberts & Summerfield, 1981) and lexically-based adaptation (e.g., Samuel, 1997, 2001; Samuel & Lieblich, 2014) are related to the different number of participants used between studies.[1] If so, then using a greater number of participants in crossmodal adaptation may reveal that adaptation can be induced in both contexts.

Experiment 2

Experiment 2 replicates the crossmodal adaptation conditions of Experiment 1, but employs larger groups of participants. If the crossmodal speech adaptation found by Roberts and Summerfield (1981) and ourselves (Experiment 1) are reliable (though small), as suggested by Meta-Analysis 1, then increasing statistical power (by employing larger groups of participants) should produce statistically significant crossmodal adaptation.

Methods

Participants. One hundred and twenty undergraduates (67 female, 53 male; M_{age} = 19.283 years, SE_{age} = .191) from the University of California, Riverside, participated in partial fulfilment of course credit. All participants were native English speaking with normal hearing and normal or corrected-to-normal vision.

Materials. The same materials used in Experiment 1 were employed in Experiment 2.

Procedure. The general procedure for Experiment 2 was the same as for Experiment 1. However, participants were split up evenly among four groups, consisting of only the crossmodal conditions described in Experiment 1: 1) auditory test-continuum and visual-/ba/ adaptor; 2) auditory test continuum and visual-/va/ adaptor; 3) visual test continuum and auditory-/ba/ adaptor; and 4) visual test continuum and auditory-/va/ adaptor. Each group thus consisted of 30 participants, consistent with selective adaptation studies investigating more subtle effects (e.g., Samuel & Lieblich, 2014; Vroomen & Baart, 2009; Vroomen et al., 2007). As with the Experiment 1 groups, half of the participants listened to tokens derived from Speaker S1 and the other half listed to tokens derived from Speaker S2.

Results & Discussion

Participant identification responses were coded using the same procedure from Experiment 1 (*see Figure 3*). Again, each experimental group was evaluated to determine whether the phonemic category boundary significantly changed after adaptation, using one-tailed paired-samples t-tests.

As illustrated in *Table 2.4*, we find that using larger groups of participants resulted in some statistically significant phonemic boundary shifts following crossmodal selective speech adaptation. Specifically, we find that adaptation to /ba/ can produce speech perception aftereffects across modalities, whether using an auditory adaptor and testing on a visual continuum or a visual adaptor and testing on an auditory continuum.

Selective adaptation to /va/ failed to produce any significant changes following crossmodal adaptation.

For both syllables, the magnitude of changes to phonemic boundaries did not differ between conditions where perceivers were adapted to auditory information and tested on visual or when adapted to visual information and tested on auditory. The significant change in auditory perception following adaptation to visual-/ba/ did not differ from the significant change in visual perception following adaptation to auditory-/ba/, $t(58) = 0.645$, $p = .261$, $r = .084$. Similarly, the non-significant change in auditory perception following adaptation to visual-/va/ did not differ from the non-significant change in visual perception following adaptation to auditory-/va/, $t(58) = -0.413$, $p = .341$, $r = .054$.

The results suggest that speech identification can change following crossmodal selective speech adaptation, even if only for one of the two syllables employed in the current investigation. We contemplate why we find changes following adaptation to only one syllable, and why crossmodal adaptation is so much smaller than within-modality adaptation (Experiment 1), in the General Discussion.

Meta-Analysis 2

The results of Experiment 2 provide both significant and non-significant crossmodal adaptation. Finding that some phonetic information (/va/) does not adapt across the auditory and visual modalities may have an impact on the mean crossmodal adaptation calculated across studies in Meta-Analysis 1, perhaps even finding the new mean to be non-significant. To determine if the results of Experiment 2 impact the mean crossmodal adaptation observed in Meta-Analysis 1, we added the four crossmodal conditions from Experiment 2 to the crossmodal conditions tested in Experiment 1 and those tested by Roberts and Summerfield (1981) in Meta-Analysis 2. This second meta-analysis was executed using the same procedure for Meta-Analysis 1.

Results & Discussion

Table 2.3 reports the results of Meta-Analysis 2. The results reveal that, across studies and conditions (auditory-to-visual, visual-to-auditory), the mean phonemic boundary shift following crossmodal adaptation (*M* = 0.123, *SE* = 0.032) is still significantly greater than 0, *t*(172) = 3.126, *p* < .001, *r* = .246. A test of heterogeneity revealed that the phonemic boundary shifts observed did not significantly differ between studies and conditions, $\chi^2(9) = 7.451$, *p* = .590.

An important thing to notice between Meta-Analysis 1 and Meta-Analysis 2 is the small change in overall mean and effect-size for crossmodal adaptation. These results suggest that the mean crossmodal adaptation observed across studies is reliable and not substantially influenced by the addition of new studies, even when (within individual studies) some types of phonetic information fail to adapt (i.e., /va/, Experiment 2). In other words, the degree to which crossmodal adaptation can change speech perception is relatively fixed. Future studies can be generally predicted to produce crossmodal adaptation similar to the mean calculated from Meta-Analysis 2.

Experiment 3

A question that arises from the results of the previous experiments and meta-analyses regards what is the basis of crossmodal selective speech adaptation and why crossmodal adaptation produces changes in speech identification that are substantially smaller than changes following within-modality adaptation.

It may be that within-modality and crossmodal adaptation may depend on common modality-neutral information. For example, the articulations of the vocal tract shape speech sounds while producing simultaneous and correlated changes in visible face structure. As such, auditory (heard) speech shares a lawful relationship with the visible (lipread) speech associated with it (e.g., Fowler, 2004; Rosenblum, 2008). Finding that crossmodal adaptation produces changes in speech identification that are smaller than changes following adaptation to within-modality information may simply be the result of the *amount* of information overlap within and between modalities (e.g., Rosenblum et al., 2007). A reasonable assumption to make would be that there is more information overlap between adaptors and test-stimuli of the same modality than between adaptors and test-stimuli of different modalities. If less information overlaps across modalities, then crossmodal adaptation would be *expected* to produce smaller changes in speech identification compared to adaptation and testing in the same modality.

Alternatively, changes in speech identification following crossmodal adaptation may have a completely different basis than when adapting and testing in the same modality. It may be that the phonetic associations learned between heard and seen speech cues (e.g., Diehl, Lotto, & Holt, 2004; Massaro, 2015) may affect selective speech adaptation. Text has previously been used as a tool for determining whether audio-visual influences in speech are due (at least in part) to the lawful relationship between heard and seen speech articulations or to the associations learned between heard and seen speech cues (e.g., Fowler & Dekle, 1991; Massaro, Cohen, & Thompson, 1988; Saldaña & Rosenblum, 1993). Text is associated with phonetics through learned convention, but does not share any lawful relationship with heard and seen speech articulations.

Previously, Cooper (1975) described an unpublished study investigating the changes in auditory speech identification that followed adaptation to text-words. He hypothesized that if selective speech adaptation depends on the phonetic associations learned between heard and seen speech cues, then adaptation to text should change auditory speech identification. Though details are not provided (Cooper, 1975), Cooper (1979) later described the results: "Although some adaptation effects were obtained using this procedure, they exhibited directional asymmetries and inconsistencies not mirrored by the typical speech adaptation effects," (Cooper, 1979, p.181). Since Cooper (1975; 1979) is vague on what changes in auditory speech identification followed text-adaptation (and the data does not appear to have ever been formally published), we decided to test Cooper's (1975) hypothesis using text-adaptors of our tested syllables. As such, Experiment 3 evaluates whether repeated presentation of visible text ("ba" or "va") could produce the same changes in auditory and visual speech identification observed in Experiment 2.

Methods

Participants. One-hundred and nineteen undergraduates (69 female, 50 male; M_{age} = 19.706 years, SE_{age} = .216) from the University of California, Riverside, participated in partial fulfillment of course credit. All participants were native English speaking with normal hearing and normal or corrected-to-normal vision.

Materials. The same materials used in Experiments 1 and 2 were employed in Experiment 3. However, the auditory and visual adaptors were replaced with text. Orthographic characters, either “ba” or “va”, were presented on a computer screen in place of video or sound adaptors at font size of 100. These text adaptors were presented following the same procedures for the auditory and visual adaptors used in Experiment 1 and 2, flashing on the screen for the same amount of time (and at the same rate) that the auditory and visual adaptors were presented.

Procedure. The procedure for Experiment 3 was similar to that of Experiment 2. However, participants were split up evenly among four groups adapted to text: 1) auditory test-continuum and text-“ba” adaptor; 2) auditory test-continuum and text-“va” adaptor; 3) visual test-continuum and text-“ba” adaptor; and 4) visual test-continuum and text-“va” adaptor. Consistent with Experiment 2, each group consisted of 30 participants, except for the group adapted to text-“va” and tested on the visual continuum, which had 29 participants (originally there were 30 participants, but one was dropped for not properly following the procedure). As with Experiments 1 and 2, half of the participants listened to tokens derived from Speaker S1 and the other half listed to tokens derived from Speaker S2. Participants were instructed not to read-aloud or mouth the text displayed on the screen during adaptation and were monitored throughout the experiment.

Results & Discussion

Participant identification responses were coded using the same procedure from Experiments 1 and 2 (*see Figure 4*). Again, each experimental group was evaluated to determine whether the phonemic category boundary significantly changed after adaptation, using 2-tailed paired-samples t-tests.

As illustrated in *Table 2.5*, we did not find changes in auditory or visual speech identification following adaptation to text-“ba”. We did find differences in auditory and visual speech identification following adaptation to text-“va”, however, in the *opposite direction* from that expected for selective speech adaptation. Perceivers identified *more* visual test-stimuli and (marginally) more auditory test stimuli (p = .102) as “va” following adaptation to text-“va”.

Changes of this nature, which are opposite to changes expected from selective speech adaptation, may demonstrate a different phenomenon, such as priming or *phonetic recalibration*. Phonetic recalibration refers to the perceptual learning that occurs when adapted to an ambiguous speech syllable in combination with some disambiguating stimulus. Following adaptation, ambiguous speech syllables are identified consistent with the disambiguating stimulus (for a review, see Vroomen & Baart, 2012). For example, perceivers will identify more ambiguous stimuli along an /aba/-to-/ada/ auditory test continuum as /aba/ following adaptation to an ambiguous auditory utterance (heard half of the time as /aba/ and half of the time as /ada/) dubbed onto a video of a face clearly articulating /aba/ (e.g., Bertelson et al., 2003). Phonetic recalibration is believed to be a result of the perceptual system learning to identify ambiguous speech information relative to the disambiguating context in which that information is experienced (e.g., Vroomen & Baart, 2012).

Interestingly, Keetels, Schakel, Bonte, and Vroomen (2016) recently discovered that adaptation to ambiguous auditory information (between /aba/ and /ada/) disambiguated by text (“aba” or “ada”) can induce phonetic recalibration. As with other examples, they explain text-induced recalibration as evidence of a perceptual system that has learned to identify ambiguous auditory speech relative to the disambiguating context (text) in which that ambiguous speech was previously experienced. However, our Experiment 3 results suggest that text need not be paired with ambiguous auditory speech to induce recalibration, at least for /va/ information. As such, recalibration from text may not be the result of a learned association between ambiguous auditory speech and read text. Instead, the recalibration observed by Keetels et al. (2016) may be explained by some other mechanism. Perhaps text primed perceivers to identify ambiguous acoustic stimuli. Future work in the area of phonetic recalibration will need to account for text-induced recalibration without a paired ambiguous auditory stimulus.

Regardless, the results of Experiment 3 suggest that text cannot induce selective speech adaptation. The results support the notion that the crossmodal adaptation observed in Experiment 2 results from the lawful relationship between auditory and visual speech.

General Discussion

Despite the non-significant crossmodal adaptation reported by Roberts and Summerfield (1981), the results of this investigation suggest that at least some phonetic information can adapt across sensory modalities. In fact, the meta-analyses that we report may suggest that the small crossmodal adaptation in the data of Roberts and Summerfield (1981) may, in fact, have been reliable, and that the failure to reach statistical significance was a result of low statistical power from a small sample size. The meta-analyses suggest, and Experiment 2 supports, the notion that though crossmodal speech adaptation is small, it is reliable. Next we will discuss specific aspects of our results and conclude by addressing the theoretical implications of crossmodal adaptation.

Segment differences in crossmodal selective speech adaptation

We can only speculate why we observe crossmodal changes in speech identification following adaptation to /ba/ but not following adaptation to /va/. The literature on selective speech adaptation suggests there is variability between different types of phonetic information to induce adapted changes in speech perception (e.g., Cooper et al., 1976; Diehl, 1975; Ganong, 1978; Saldaña & Rosenblum, 1994; Samuel, 1997; Samuel & Lieblich, 2014). For example, Diehl (1975) reported means suggesting that changes in the identification of test-stimuli along an auditory /bɛ/-to-/dɛ/ continuum following adaptation to auditory-/bɛ/ were twice that following adaptation to auditory-/dɛ/ (see also Roberts & Summerfield, 1981). In contrast, Samuel (1997) found that changes in the identification of test-stimuli along an auditory /bI/-to-/dI/ continuum following adaptation to auditory word-utterances containing a critical /d/ segment were about twice that following adaptation to auditory word-utterances containing a critical /b/ segment (this difference was mirrored when adapted to restored words). Similarly, Saldaña and Rosenblum (1994) report that changes in the identification of test-stimuli along an auditory /ba/-to-/va/ continuum following adaptation to auditory-/ba/ were significantly greater than changes following adaptation to auditory-/va/ (though both adaptors induced significant change). In other words, perceivers identified more auditory test-stimuli as /va/ following adaptation to auditory-/ba/ than they identified as /ba/ following adaptation to auditory-/va/. However, we do not observe the same pattern of results in Experiment 1 (see Table 2.1). We instead find no significant difference in magnitude of change following adaptation to auditory-/ba/ and auditory-/va/ (*p* = .188). Variability across studies for how particular phonetic segments induce changes in speech identification following adaptation could be attributed to the subtle differences in how adaptors and test-stimuli are constructed.

What is of particular interest to the current investigation is why we observe bidirectional adaptation when adapting and testing in the same modality (Experiment 1) but only observe crossmodal adaptation for /ba/ (Experiment 2). The basis for this directional asymmetry may lie with the technique used to construct our stimuli. Interestingly, Cooper et al. (1976) observed an analogous asymmetry in their test of adaptation between auditory speech perception and speech production: Perceivers failed to demonstrate a change in categorization of speech along /si/-to-/sti/ and /su/-to-/stu/ auditory continua following repeated production of /si/ or /su/, though changes did follow repeated production of /sti/ and /stu/. However, when adapted to auditory utterances of /si/ and /su/, and /sti/ and /stu/, perceivers exhibited changes in both directions. Cooper et al. (1976) offered a very speculative suggestion that perhaps their unidirectional effects have something to do with the liquid nature of /si/-/su/ productions that differed structurally from the /si/-to-/sti/ and /su/-to-/stu/ test stimuli. Their test continua were constructed by taking utterances of /si/ and /su/ and gradually added silence (in 10ms increments) between the /s/ and /i/ or /u/. Their finished continua comprised stimuli that had at least a 10ms of silence between the /s/ and the /i/ or /u/, even at the /si/ and /su/ ends of their continua; 10ms of silence that is not present when producing /si/ and /su/. If the difference between the productions of perceivers and the test continua accounts for their unidirectional effects, Cooper et al. (1976) suggested that adaptation may depend on the presence of shared acoustic properties between speech production and auditory test-syllables.

The pattern of our results is similar to those observed by Cooper et al. (1976). Though we observe bidirectional effects when participants were adapted and tested in the same modality, either auditory or visual (Experiment 1), we only find a crossmodal effect when adapting /ba/ (Experiment 2). It may be that some types of phonetic information important to selective adaptation do not cross modalities. Alternatively, the differences we find between within-modality and crossmodal adaptation could have something to do with the procedure used to construct our stimuli. Remember that our auditory /ba/ and /va/ stimuli were derived from the same /va/ utterance, similar to how Cooper et al. (1976) constructed their /si/-/sti/ and /su/-/stu/ test-continua. At this point, we cannot rule out either of these possibilities.

Differences between within and crossmodal speech adaptation

We found in Experiment 1 that within-modality speech adaptation is substantially greater than crossmodal adaptation. Remember that Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994) reported that, when employing bimodal adaptors, selective adaptation to within-modality information dominates over competing crossmodal information. The evidence seems to suggest that though speech information may adapt across sensory modalities, sensory-consistent information, if given the chance, will adapt first, or at least more strongly. This raises questions regarding what information is adapted at the sensory level and whether this differs from the information adapted across modalities.

Other studies have used selective speech adaptation to investigate whether different levels of processing are sensitive to different forms of information. Interestingly, these studies have generally suggested that peripheral (sensory) processes are more sensitive to the spectrotemporal overlap between adaptors and test-stimuli. For example, several studies have demonstrated an ipsilateral advantage when testing interaural differences in selective speech adaptation (Ades, 1974; Jamieson & Cheesman, 1986; Samuel & Kat, 1996; Sawusch, 1977). Ades (1974) found that adaptation to speech in one ear can change auditory speech identification in the other ear. However, the change in auditory speech identification in one ear following adaptation in the opposite ear was only about 56% of the change observed when adapted and tested in the same ear. Based on this evidence, Ades (1974) proposed two levels of processing during selective speech adaptation, one level peripheral and another central. In a follow-up investigation, Sawusch (1977) found that adaptation and testing within the same ear is strongly influenced by the spectral overlap between auditory adaptors and auditory test stimuli. However, adaptation across ears is more flexible and can be similarly influenced across a winder frequency spectrum (while preserving the phonetic structure between adaptors and test stimuli). Based on these and their own findings, Samuel and Kat (1996) later suggested that selective speech adaptation is sensitive to patterns of acoustic energy at peripheral levels of processing and to more complex (phonetic) forms of information at higher (central) levels of processing.

One possible explanation for the degree of difference between within and crossmodal speech adaptation may be that low-level sensory information pertaining to the acoustic or visual qualities of the stimulus (e.g., tone, frequency, intensity, etc.) adapt first, rather than speech (phonetic) information as such (e.g., Kat & Samuel, 1984; Roberts & Summerfield, 1981). However, in the absence of shared sensory information between an adaptor and test-stimuli, adaptation relies on common information preserved across the different energy media (light and sound) (e.g., Fowler, 2004; Rosenblum, 2005). This information is related directly to the distal articulatory events that cause simultaneous and correlated changes in the structure of energy media (e.g., Summerfield, 1987). Since different information may affect adaptation within and across sensory modalities, different processes may be at play, which might account for why crossmodal adaptation is weaker.

Alternatively, selective adaptation within and across modalities may *both* be based on lawful articulatory information affecting a common mechanism, without invoking multiple levels of processing. As we suggested previously, changes in speech identification when adapting and testing in the same modality may be greater than changes when adapting and testing in different modalities because there is more overlapping articulatory information when adapting and testing in the same modality (e.g., Rosenblum et al., 2007). The notion that crossmodal speech adaptation is affected by common lawfully related (articulatory) information available across the auditory and visual modalities is supported by our Experiment 3 results. Text is not lawfully related to articulation in the same way as audible and visible speech and the changes in auditory or visual speech identification that followed text adaptation did not reflect selective speech adaptation, as such.

Tangentially, the Experiment 3 results also suggests that the observed changes in speech identification following crossmodal adaptation are unlikely to be explained by some change in perceiver decision criterion (e.g., Storrs, 2015; Storrs & Arnold, 2012). For example, if crossmodal adaptors induced changes in speech identification by (consciously or non-consciously) affecting the cognitive strategy employed by the perceiver (e.g., "I see /va/, so I should identify more stimuli as opposite to /va/".), then the same strategy for test-stimulus identification would be expected to follow text adaptation. Instead, text adaptation resulted in either no adaptation or changes in speech identification opposite to those following crossmodal adaptation.

Remember too that the participants in Experiment 3 were instructed (and monitored) not to read-aloud or mouth the text adaptors during adaptation. However, they could still have subvocalized the text during adaptation. Based on the result of Experiment 3, it is unlikely that the crossmodal adaptation observed in Experiment 2 is the result of subvocalization of visible speech information.

Implications for theoretical accounts

Theories explaining selective speech adaptation will need to account for the crossmodal influences observed in the current investigation (e.g., Dias et al., 2016; Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014).

For example, Roberts and Summerfield (1981) suggested that selective speech adaptation depends on sense-specific spectrotemporal overlap between adaptors and test-stimuli. However, the current investigation suggests that (subtle) crossmodal adaptation it is reliable, at least for some forms of phonetic information.

Recall also that Roberts and Summerfield (1981), and later Saldaña and Rosenblum (1994), found changes in auditory speech identification following adaptation to audio-visual incongruent adaptors (e.g., auditory-/bɛ/-visual-/gɛ/, perceived as /dɛ/) (e.g., McGurk & MacDonald, 1976). These changes reflected adaptation to the auditory component of the bimodal adaptors, not the integrated audio-visual phonetic percepts. Based on these results, and the broader literature demonstrating multisensory influences in speech, Saldaña and Rosenblum (1994) proposed that selective speech adaptation may occur at a level of processing prior to crossmodal integration of auditory and visual speech. However, finding subtle crossmodal adaptation in the current investigation allows for a new interpretation. We suggested earlier that changes in speech perception are greater following within-modality adaptation than following crossmodal adaptation because more articulatory information is shared between adaptors and test-stimuli of the same modality than between adaptors and test-stimuli of different modalities. It may be that Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994) observed changes in auditory speech identification that reflected adaptation to the auditory components of their audio-visual incongruent adaptors because the auditory components shared more overlapping information with the auditory test-stimuli than the visual components (or the integrated audio-visual percepts). A similar explanation could be applied to other studies investigating the role of bimodal speech information in selective speech adaptation (e.g., Dias et al., 2016; Samuel & Lieblich, 2014).

Though the results of the current investigation suggest that selective speech adaptation is sensitive to modality neutral articulatory information, the role of lexical knowledge in selective speech adaptation requires consideration. Restored-word adaptors (i.e., the phonemic restoration effect) do not share lawfully generated articulatory information with test-stimuli (e.g., Samuel, 1997, 2001; Samuel & Lieblich, 2014). Lexical knowledge (i.e., words) is associated with phonetics though linguistic experience; a learned association. The illusory syllables perceived within the context of restored-words result from the top-down influence of lexical knowledge (e.g., Samuel, 1981, 1987). However, adaptation to these illusory syllables can change (auditory) speech perception (e.g., Samuel, 1997, 2001).

Different mechanisms likely account for changes in speech perception following crossmodal adaptation and following adaptation to lexically influenced phonetic percepts. In fact, the initial premise of the theoretical account offered by Samuel and Lieblich (2014) may be correct: Selective speech adaptation may be first influenced by bottom-up information and only in the absence of bottom-up information can top-down (lexical/linguistic) influences affect selective speech adaptation. However, based on the results of the current investigation, bottom-up information can be modality-neutral (e.g., Gibson, 1966). The subsequent premises of the theoretical account offered by Samuel and Lieblich (2014), describing the perceptual influences of visual speech as top-down and non-linguistic to account for the lack of crossmodal influences observed in previous studies (e.g., Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994), are likely invalid. Though future studies will need to consider why the top-down associations of words to phonetics can induce selective speech adaptation while the top-down associations of text to phonetics cannot (Experiment 3).

Crossmodal aftereffects in speech also have implications for the general theoretical understanding of the speech process. Theorists who have previously cited the non-significant crossmodal adaptation reported by Roberts and Summerfield (1981) to support theories proposing a perceptual mechanism that treats auditory and visual speech information differently will need to account for crossmodal influences in selective speech adaptation (e.g., Diehl et al., 1985; Massaro, 1987; Samuel & Lieblich, 2014). In fact, the results seem consistent with evidence for a speech form that is accessible across sensory modalities at the earlies stages of perceptual processing (for reviews, see Fowler, 2004; Rosenblum, 2005; Rosenblum et al., 2016). Beyond this, the results are consistent with audio-visual crossmodal adaptation observed outside of speech (e.g., Ehrenstein & Reinhardt-Rutland, 1996; Kitagawa & Ichihara, 2002; Levitan et al., 2015), and add to a growing body of literature suggesting a general perceptual mechanism that is amodal, instead concerned with the acquisition of information that is available across sensory modalities (Lacey, Campbell, & Sathian, 2007; Rosenblum et al., 2016; Shams & Kim, 2010).

References

Ades, A. E. (1974). Bilateral component in speech perception? *Journal of the Acoustical Society of America, 56*(2), 610-616.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92*(2), 339-355.

Baart, M., & Vroomen, J. (2010). Do you see what you are hearing? Cross-modal effects of speech sounds on lipreading. *Neuroscience Letters, 471,* 100-103.

Bernstein, L. E., & Liebenthal, E. (2014). Neural pathways for visual speech perception. *Frontiers in Neuroscience, 8*(386). doi:10.3389/fnins.2014.00386

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science, 14*(6), 592-597.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.53). Retrieved from http://www.praat.org/

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instruments and Computers, 25*(2), 257-271.

Cooper, W. E. (1974). Perceptuomotor adaptation to a speech feature. *Perception & Psychophysics, 16*(2), 229-234.

Cooper, W. E. (1975). Selective Adaptation to Speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive Theory* (Vol. 1, pp. 23-54). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooper, W. E. (1979). *Speech Perception and Production: Studies in Selective Adaptation*. Norwood, NJ: Ablex Publishing Corporation.

Cooper, W. E., Billings, D., & Cole, R. A. (1976). Articulatory effects on speech perception: A second report. *Journal of Phonetics, 4*(3), 219-232.

Cooper, W. E., Blumstein, S. E., & Nigro, G. (1975). Articulatory effects on speech perception: A preliminary report. *Journal of Phonetics, 3,* 87-98.

Cooper, W. E., & Lauritsen, M. R. (1974). Feature processing in the percption and production of speech. *Nature, 252*, 121-123.

Cumming, G. (2012). *Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York, NY: Taylor & Francis Group, LLC.

Derrick, D., & Gick, B. (2013). Aerotactile Integration from Distal Skin Stimuli. *Multisensory research, 26*(5), 405-416.

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. doi:10.1016/j.wocn.2016.02.004

Diehl, R. L. (1975). The effect of selective adaptation on the identification of speech sounds. *Perception & Psychophysics, 17*(1), 48-52.

Diehl, R. L., Kluender, K. R., & Parker, E. M. (1985). Are selective adaptation and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception and Performance*, *11*(2), 209-220.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Carnegie Mellon University Research Showcase*. Retrieved from http://repository.cmu.edu/psychology/155

Ehrenstein, W. H., & Reinhardt-Rutland, A. H. (1996). A cross-modal aftereffect: Auditory displacement following adaptation to visual motion. *Perception and Motor Skills*, *82*, 23-26.

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

Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders, 40*(4), 481-492.

Finney, D. J. (1971). *Probit Analysis*. Cambridge, MA: Cambridge University Press.

Fowler, C. A. (2004). Speech as a supramodal or amodal phenomenon. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processing* (pp. 189-202). Cambridge, MA: MIT Press.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17*(3), 816-828.

Ganong, W. F. (1978). The selective adaptation effects of burst-cued stops. *Perception & Psychophysics, 24*(1), 71-83.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin Company.

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature, 426*, 502-504.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251-279.

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.

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *PNAS, 106*(4), 1245-1248.

Jamieson, D. G., & Cheesman, M. F. (1986). Locus of selective adapation in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 12*(3), 286-294.

Jones, B. C., Feinberg, D. R., Bestelmeyer, P. E. G., DeBruine, L. M., & Little, A. C. (2010). Adaptation to different mouth shapes influences visual perception of ambiguous lip speech. *Psychonomic Bulletin & Review, 17*(4), 522-528.

Kat, D., & Samuel, A. G. (1984). More adaptation of speech by nonspeech. *Journal of Experimental Psychology: Human Perception and Performance, 10*(4), 512-525.

Keetels, M., Schakel, L., Bonte, M., & Vroomen, J. (2016). Phonetic recalibration of speech by text. *Attention, Perception, & Psychophysics, 78*(3), 938-945. doi:10.3758/s13414-015-1034-y

Kim, J., Davis, C., & Krins, P. (2004). Amodal processing of visual speech as revealed by priming. *Cognition, 93*(1), B39-B47. doi:http://dx.doi.org/10.1016/j.cognition.2003.11.003

Kitagawa, N., & Ichihara, S. (2002). Hearing visual motion in depth. *Nature, 416*, 172-174.

Lacey, S., Campbell, C., & Sathian, K. (2007). Vision and touch: Multiple or multisensory representations of objects. *Perception, 36*, 1513-1521.

Levitan, C. A., Ban, Y.-H. A., Stiles, N. R. B., & Shimojo, S. (2015). Rate perception adapts across the senses: evidence for a unified timing mechanism. *Sci. Rep., 5*. doi:10.1038/srep08857

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (2015). Speech perception. In J. D. Wright (Ed.), *International Encyclopedia of Social & Behavioral Sciences* (2nd ed., Vol. 23, pp. 235-242). Oxford: Elsevier.

Massaro, D. W., Cohen, M. M., & Thompson, L. A. (1988). Visible language in speech perception: Lipreading and reading. *Visible Language, 22*(1), 8-31.

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

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics, 72*(6), 1614-1625.

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research, 71*, 4-12.

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. doi:http://dx.doi.org/10.1016/j.cognition.2016.02.019

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, *119*(4), 2382-2393.

Pilling, M., & Thomas, S. (2011). Audiovisual cues and perceptual learning of spectrally distorted speech. *Language and Speech, 54*, 487-497.

Pisoni, D., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, *25*, 21-52.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lip-Reading* (pp. 97-113). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Remez, R. E., Fellowes, J. M., Pisoni, D. B., Goh, W. D., & Rubin, P. E. (1998). Multimodal perceptual organization of speech: Evidence from tone analogs of spoken utterances. *Speech Communication*, *26*, 65-73.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics, 30*(4), 309-314.

Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. Pisoni & R. Remez (Eds.), *Handbook of Speech Perception* (pp. 51-78). Malden: Blackwell.

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science, 17*(6), 405-409.

Rosenblum, L. D., Dias, J. W., & Dorsi, J. J. (2016). The supramodal brain: implications for auditory perception. *Journal of Cognitive Psychology, 2016*. doi:10.1080/20445911.2016.1181691

Rosenblum, L. D., Miller, R. M., & Sanchez, K. (2007). Lip-read me now, hear me better later: Cross-modal transfer of talker-familiarity effects. *Psychological Science, 18*(5), 392-396.

Rosenblum, L. D., & Saldaña, H. M. (1992). Discrimination tests of visually influenced syllables. *Perception and Psychophysics, 52*(4), 461-473.

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 22*(2), 318-331.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17*, 1147-1153.

Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics*, *54*(3), 406-416.

Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *Journal of the Acoustical Society of America, 95*(6), 3658-3661.

Sams, M., Mottonen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research, 23*, 429-435.

Samuel, A. G. (1981). Phonemic Restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*(4), 474-494.

Samuel, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Langauge, 26*, 36-56.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32*, 97-127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*(4), 348-351.

Samuel, A. G., & Kat, D. (1996). Early levels of analysis of speech. *Journal of Experimental Psychology: Human Perception and Performance, 22*(3), 676-694.

Samuel, A. G., & Kat, D. (1998). Adaptation is Automatic. *Perception & Psychophysics, 60*(3), 503-510.

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), 1479-1490.

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.

Sanchez, K., Dias, J. W., & Rosenblum, L. D. (2013). Experience with a talker can transfer across modalities to facilitate lipreading. *Attention, Perception, & Psychophysics, 75*(7), 1359-1365.

Sawusch, J. R. (1977). Peripheral and central processes in selective adaptation of place of articulation in stop consonants. *Journal of the Acoustical Society of America, 62*(3), 738-750.

Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews, 7*, 269-284.

Storrs, K. R. (2015). Are high-level aftereffects perceptual? *Frontiers in Psychology, 6*, 157. doi:10.3389/fpsyg.2015.00157

Storrs, K. R., & Arnold, D. H. (2012). Not all face aftereffects are equal. *Vision Research, 64*, 7-16. doi:http://dx.doi.org/10.1016/j.visres.2012.04.020

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*(4), 661-699.

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

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In I. Barbara & C. Ruth (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3-51). Hillsdale, NJ: Erlbaum.

Vroomen, J., & Baart, M. (2009). Phonetic recalibration only occurs in speech mode. *Cognition, 110*, 254-259.

Vroomen, J., & Baart, M. (2012). Phonetic recalibration in audiovisual speech. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes* (pp. 363-379). Boca Raton, FL: CRC Press.

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*, 572-577.

Footnotes

[1] Interestingly, there are similar patterns of small change in auditory speech identification following adaptation to restored-/ba/ words (Samuel, 1997) and following adaptation to visual-/ba/ (Experiment 1 results). Samuel (1997) measured phonemic boundary shifts as the difference in the identification of the middle three tokens of his 8-step /ba/-to-/da/ auditory continuum before and after adaptation. Using this technique, Samuel (1997) found that these middle three tokens were 2.1% less likely to be identified as /ba/ follow adaptation to restored /ba/-words. When we calculated shifts from our own data using this same technique (from the middle three tokens of our 11-step /va/-to-/ba/ continuum), we find that the middle three tokens are 4.4% less likely to be identified as /ba/ following adaptation to visual-/ba/.

Table 2.1

Experiment 1: Phonemic category boundaries and magnitudes of boundary shift following adaptation.

Continuum Modality	Adaptor Modality	Adaptor	Phonemic Boundary		Shift	*SE*	*t*	*n*	$p_{(1\text{-}tailed)}$		$r_{effect\ size}$
			Baseline	Adapted							
Auditory	Auditory	/ba/	4.993	6.342	1.350	0.313	4.317	10	0.001	***	0.821
		/va/	5.664	3.827	-1.837	0.436	-4.212	10	0.001	***	0.815
	Visual	/ba/	5.945	6.098	0.153	0.171	0.899	10	0.196		0.287
		/va/	6.094	5.649	-0.445	0.450	-0.989	10	0.175		0.313
Visual	Auditory	/ba/	5.774	5.996	0.222	0.165	1.344	10	0.106		0.409
		/va/	5.937	5.957	0.020	0.206	0.097	10	0.463		-0.032
	Visual	/ba/	5.901	6.579	0.678	0.300	2.261	10	0.025	*	0.602
		/va/	5.877	5.137	-0.740	0.155	-4.771	10	0.001	***	0.847

Notes: *p < .05, ***p < .001. Negative r-values indicate counter-hypothetical boundary shifts.

Table 2.2

Experiment 1: Differences between the phonemic boundary shifts following within-modality and crossmodal adaptation.

Continuum	Within-Modality	Crossmodal	Difference in Phonemic						
Modality	Adaptor	Adaptor	Boundary Shift	*SE*	*t*	*n*	$p_{(1\text{-tailed})}$		$r_{effect\ size}$
Auditory	Auditory-/ba/	Visual-/ba/	1.196	0.356	3.360	20	0.002	**	0.621
	Auditory-/va/	Visual-/va/	-1.392	0.627	-2.220	20	0.020	*	0.464
Visual	Visual-/ba/	Auditory-/ba/	0.456	0.342	1.332	20	0.100	†	0.300
	Visual-/va/	Auditory-/ba/	-0.760	0.258	-2.943	20	0.005	**	0.570

Notes: †*p* < .10 (marginal significance), **p* < .05, ***p* < .01. Differences in phonetic boundary shift are calculated from the means reported in Table 2.1.

Table 2.3

Meta-analyses of crossmodal selective adaptation in speech.

	Test	Test	Adaptor									
Study Information	Modality	Continuum	Modality	Adaptor	M_{shift}	SE	t	n	$p_{(1\text{-}tailed)}$	$r_{effect\ size}$	Condition Weights	
Roberts & Summerfield (1981)	AO	/da-ba/	VO	/ba/	0.103	0.242	0.426	6	0.688	0.187	12%	3%
				/da/	0.113	0.178	0.633	6	0.554	0.273	21%	5%
Experiment 1	AO	/va-ba/	VO	/ba/	0.153	0.171	0.898	10	0.393	0.287	23%	5%
				/va/	0.445	0.450	0.988	10	0.349	0.313	3%	1%
	VO	/va-ba/	AO	/ba/	0.222	0.165	1.345	10	0.212	0.409	25%	6%
				/va/	-0.020	0.206	-0.097	10	0.925	-0.032	16%	4%
Meta-Analysis 1					**0.138**	**0.082**	**1.681**	**52**	**0.049 ***	**0.253**	**100%**	
Experiment 2	AO	/va-ba/	VO	/ba/	0.261	0.103	2.535	30	0.009 **	0.426		15%
				/va/	0.026	0.106	0.248	30	0.403	0.046		14%
	VO	/va-ba/	AO	/ba/	0.181	0.071	2.536	30	0.009 **	0.426		31%
				/va/	-0.032	0.093	-0.344	30	0.367	-0.064		18%
Meta-Analysis 2					**0.123**	**0.039**	**3.126**	**172**	**0.001 ****	**0.246**		**100%**

Notes: *p < .05, **p < .01. Overall mean and effect size values are weighted by study. Study weights were calculated based on the inverse of the squared standard error, and the overall standard error was calculated as the square root of the summed inverse weights (e.g., Cumming, 2012).

Table 2.4

Experiment 2: Phonemic category boundaries and magnitudes of boundary shift following adaptation.

Continuum Modality	Adaptor Modality	Adaptor	Phonemic Boundary Baseline	Phonemic Boundary Adapted	Shift	*SE*	*t*	*n*	$p_{(1\text{-tailed})}$		$r_{effect\ size}$
Auditory	Visual	/ba/	5.554	5.815	0.261	0.103	2.535	30	0.009	**	0.426
		/va/	5.726	5.700	-0.026	0.106	-0.248	30	0.403		0.046
Visual	Auditory	/ba/	5.882	6.063	0.181	0.071	2.536	30	0.009	**	0.426
		/va/	5.982	6.014	0.032	0.093	0.344	30	0.367		-0.064

Notes: $^{**}p < .01$. Negative *r*-values indicate counter-hypothetical boundary shifts.

Table 2.5

Experiment 3: Phonemic category boundaries and magnitudes of boundary shift following adaptation.

Continuum	Text	Phonemic Boundary								
Modality	Adaptor	Baseline	Adapted	Shift	*SE*	*t*	*n*	$p_{(2\text{-tailed})}$		$r_{effect\ size}$
Auditory	"ba"	5.788	5.651	-0.137	0.163	-0.844	30	0.405		-0.155
	"va"	5.721	5.958	0.237	0.141	1.678	30	0.104		-0.297
Visual	"ba"	6.012	5.977	-0.034	0.124	-0.277	30	0.784		-0.051
	"va"	5.888	6.289	0.401	0.114	3.502	29	0.002	**	-0.552

Notes: ***p* < .01. Negative *r*-values indicate counter-hypothetical boundary shifts.

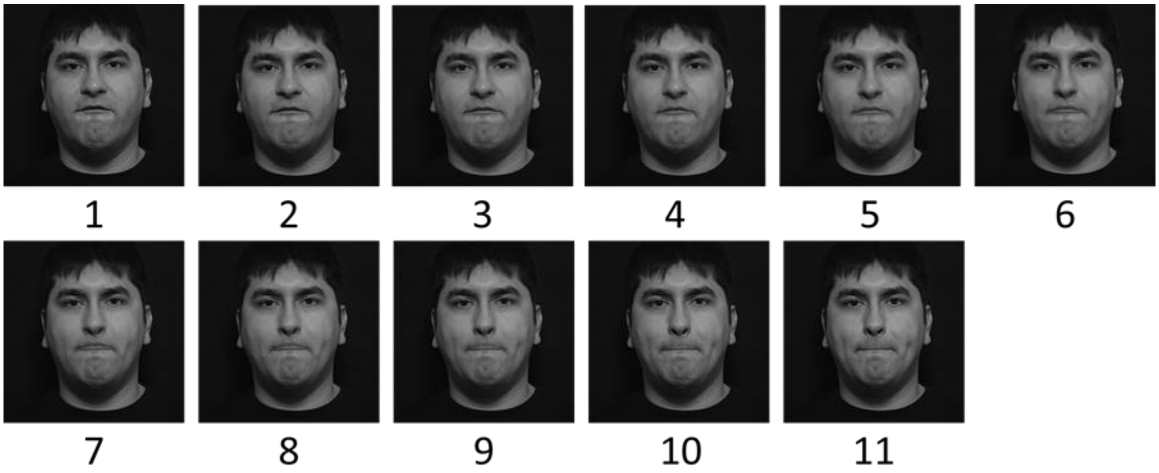


Figure 2.1. Still images take from the visual test-continuum at the place of articulation. Stimulus 1 corresponds with the unambiguous visual-/va/ video and stimulus 11 corresponds with the unambiguous visual-/ba/ video.

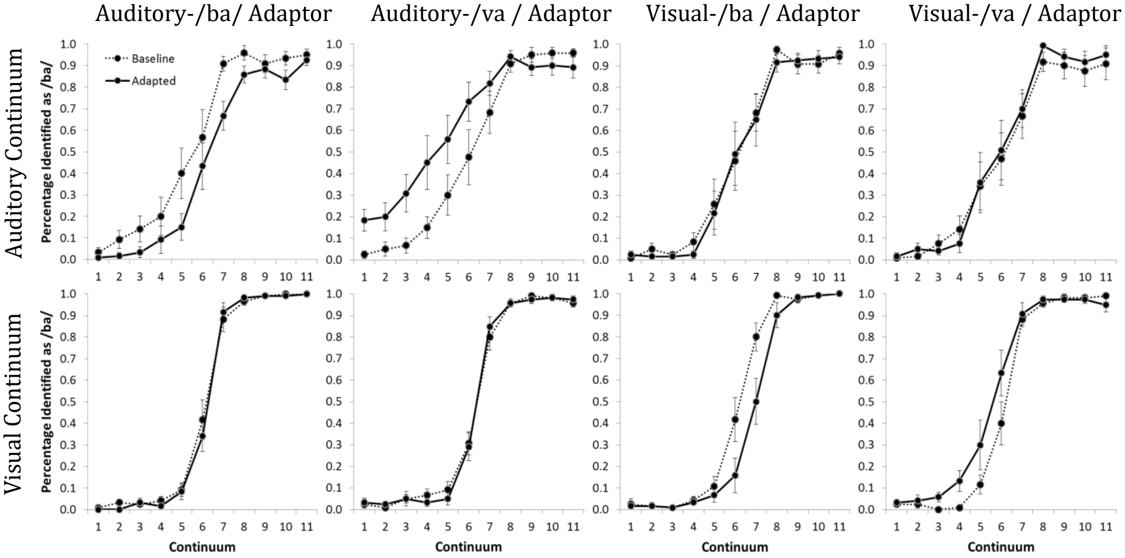
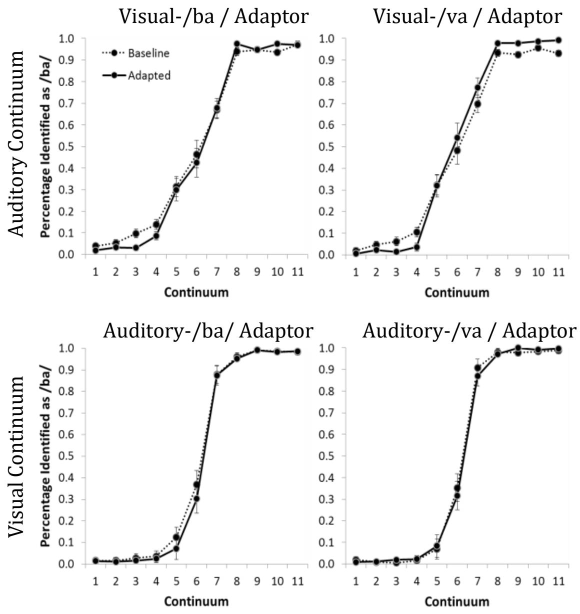


Figure 2.2. Experiment 1 test-stimulus identifications. The average proportion each continuum test-stimulus was identified as /ba/ before (Baseline, dashed line) and after (Adapted, solid line) adaptation. Data is represented for each of the eight groups, depending on tested continuum modality and adaptor. Error bars represent standard errors of the mean.


Visual-/ba / Adaptor

Visual-/va / Adaptor

Auditory-/ba/ Adaptor

Auditory-/va / Adaptor

Figure 2.3. Experiment 2 test-stimulus identifications. The average proportion each continuum test-stimulus was identified as /ba/ before (Baseline, dashed line) and after (Adapted, solid line) adaptation. Data is represented for each of the four groups, depending on tested continuum modality and adaptor. Error bars represent standard errors of the mean.

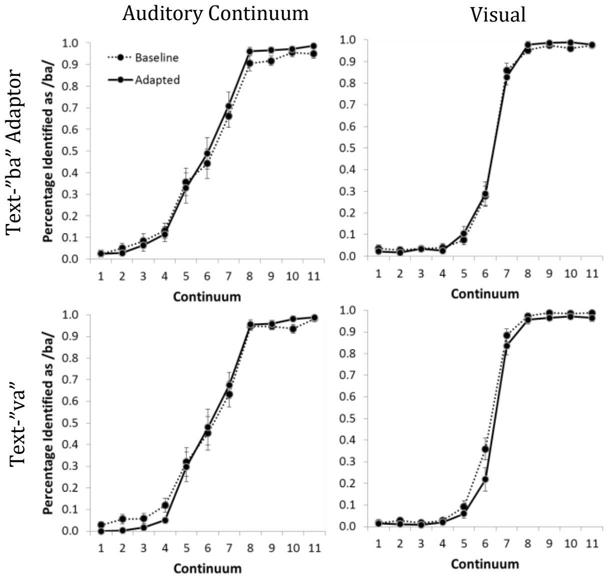


Figure 2.4. Experiment 3 test-stimulus identifications. The average proportion each continuum test-stimulus was identified as /ba/ before (Baseline, dashed line) and after (Adapted, solid line) adaptation. Data is represented for each of the four groups, depending on tested continuum modality and text adaptor. Error bars represent standard errors of the mean.

Chapter 3

Audio-visual selective speech adaptation does not exhibit a bimodal advantage

Selective speech adaptation describes when changes in speech identification follow repeated exposure (adaptation) to phonetic information (for reviews, see Cooper, 1975; Dias, Cook, & Rosenblum, 2016; Vroomen & Baart, 2012). For example, Eimas and Corbit (1973) demonstrated in their classic example how adaptation to auditory utterances of /ba/ or /pa/ can change how perceivers identify speech along an auditory /ba/-to-/pa/ continuum. Following adaptation to /ba/, perceivers identified more continuum items as /pa/, while following adaptation to /pa/, perceivers identified more continuum items as /ba/. Since their seminal study, many others have demonstrated that adaptation to auditory speech can change auditory speech identification (e.g., Ganong, 1978; Remez, 1980; Samuel & Kat, 1998) and that adaptation to visual speech can change visual speech identification (e.g., Baart & Vroomen, 2010; Dias et al., 2016; Jones, Feinberg, Bestelmeyer, DeBruine, & Little, 2010).

Studies investigating bimodal audio-visual speech adaptation suggest that sense-specific information shared between adaptors and test-stimuli will predict changes in auditory speech identification over crossmodal information (Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014). For example, Roberts and Summerfield (1981) investigated the changes in identification of test-stimuli along an auditory /bɛ/-to-/dɛ/ continuum that followed adaptation to an auditory-/bɛ/-visual-/gɛ/ bimodal adaptor, an audio-visual combination typically perceived as /dɛ/ (e.g., McGurk & MacDonald, 1976). They hypothesized that if selective adaptation is sensitive to perceived phonetic information, then more test-stimuli should be identified as /bɛ/ following adaptation. However, the changes in speech identification that they observed reflected adaptation to the auditory component of the audiovisual adaptor (more test-stimuli identified as /dɛ/). Saldaña and Rosenblum (1994) found similar effects when employing an auditory-/ba/-visual-/va/ adaptor, an audio-visual combination that produced a compelling visually influenced /va/ perception 99% of the time. These investigations suggest that sense-specific information shared between adaptors and test-stimuli drives selective speech adaptation over competing crossmodal input.

However, these examples from Roberts and Summerfield (1981) and Saldaña and Rosenblum (1994) used audio-visual adaptors that incorporate discrepant auditory and visual components. Though such audio-visual discrepant combinations can produce compelling phonetic percepts (e.g., McGurk & MacDonald, 1976), the quality of these percepts is often perceived as weaker compared to percepts derived from audio-visual stimuli with congruent components (auditory and visual components provide the same speech information) (e.g., Rosenblum & Saldaña, 1992).

In other speech domains, congruent audio-visual information can provide an advantage over audio-alone (for a review, see Rosenblum, 2008). Hearing and seeing a speaker can improve comprehension of speech spoken in noisy environments (e.g., Bernstein, Auer, Eberhardt, & Jiang, 2013; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954) and can improve comprehension of accented speech (e.g., Hardison, 2005; Navarra & Soto-Faraco, 2007; Sueyoshi & Hardison, 2005). Visual speech information can also improve comprehension of audible speech that conveys complicated content (e.g., Arnold & Hill, 2001) and can improve non-conscious imitation of another person's heard speaking style (Dias & Rosenblum, 2011, 2015). There is even evidence suggesting that visible speech information can improve auditory speech processing speed at the neural level (e.g., Paris, Kim, & Davis, 2013). Audio-visual bimodal advantages in speech processing have been described as arising from the lawful relationship between heard and seen speech events. The articulations of the vocal tract that shape speech sounds simultaneously produce visible articulations in the face. As such, auditory and visual streams provide redundant information pertaining to common articulatory events (for reviews, see Fowler, 2004; Rosenblum, 2008; Rosenblum, Dorsi, & Dias, 2016).

These examples of bimodal advantages in other speech domains have inspired some to hypothesize that changes in speech identification following bimodal speech adaptation will be greater than changes following unimodal adaptation (Dias et al., 2016; Chapter 1; Roberts & Summerfield, 1981). However, there is currently little evidence to support this hypothesis. In the same experiment described above, Roberts and Summerfield (1981) found that changes in auditory speech identification were no greater following adaptation to *congruent* audiovisual adaptors (where the auditory and visual components provide the same speech information) than changes following auditory-alone adaptation. Similarly, Dias et al. (2016; Chapter 1) later found that changes in the identification of visually-influenced audiovisual speech percepts were no greater following audiovisual adaptation than they were following visual-alone adaptation.

However, the failure of Roberts and Summerfield (1981) and Dias et al. (2016; Chapter 1) to find a significant bimodal advantage in selective speech adaptation is made somewhat complicated by recent evidence for *crossmodal* speech adaptation. Dias and Rosenblum (2016; Chapter 2) provided compelling evidence suggesting that auditory and visual syllable identification can change following adaptation to crossmodal visual and auditory syllables. In other words, adaptation to visual speech can change identification of auditory speech and adaptation to auditory speech can change identification of visual speech, at least for some types of phonetic information. Changes in speech identification following crossmodal adaptation were much more subtle compared to changes following within-modality adaptation (when adapted and tested in the same sensory modality). These subtle crossmodal effects required a much larger sample size to reach statistical significance within a single study.

Dias and Rosenblum (2016; Chapter 2) suggested that selective speech adaptation may depend (at least in part) on the overlap of modality-neutral articulatory information shared between adaptors and test stimuli. Changes in speech identification are greater following adaptation to speech in the same sensory modality because more overlapping information exists between adaptors and test-stimuli of the same modality than between adaptors and test-stimuli of different modalities (e.g., Rosenblum, Miller, & Sanchez, 2007). This conclusion was supported by the lack of perceptual change following adaptation to text-stimuli: Text is associated with phonetics by convention, but does not share any lawful relationship to heard and seen speech articulations.

Current investigation

Roberts and Summerfield (1981) and Dias et al. (2016; Chapter 1) failed to find a significant bimodal enhancement in selective speech adaptation. However, a couple of factors may have contributed to their failure.

As previously discussed, crossmodal influences in selective speech adaptation are subtle, requiring much larger sample sizes to reach statistical significance within a single study (Dias & Rosenblum, 2006; Chapter 2). Roberts and Summerfield (1981) and Dias et al. (2016; Chapter 1) employed samples of 6 and 10 participants per experimental group, respectively, when testing for bimodal enhancement of selective speech adaptation. However, Dias & Rosenblum (2016; Chapter 2) used samples of 30 participants per experimental group to find significant *crossmodal* adaptation. The subtle nature of crossmodal adaptation may subsequently render any bimodal enhancement of selective speech adaptation also subtle, requiring larger samples of participants to provide the statistical power needed to reach statistical significance (e.g., Dias & Rosenblum, 2016; Chapter 2). In fact, there may be some evidence suggesting that changes in speech identification are subtly greater following bimodal adaptation than following unimodal adaptation. Evaluating the differences in unimodal and bimodal adaptation observed by Roberts and Summerfield (1981) and Dias et al. (2016), we find that two of the three conditions tested across these studies exhibit (non-significant) differences that may suggest subtle bimodal advantages (see *Table 3.1*).

We employed larger groups of participants in the current investigation to determine whether subtle bimodal enhancement of selective speech adaptation can reach statistical significance when greater statistical power is provided by employing larger samples of participants (e.g., Dias & Rosenblum, 2016; Chapter 2).

We also evaluated whether bimodal advantages in selective speech adaptation depend on the test-modality. Roberts and Summerfield (1981) evaluated whether changes in auditory speech identification were greater following audiovisual speech adaptation than following auditory speech adaptation. Dias et al. (2016; Chapter 1) evaluated whether changes in the identification of visually-influenced audiovisual speech percepts were greater following audiovisual speech adaptation than following visual speech adaptation. However, what has not been evaluated is whether changes in visual speech identification are greater following audiovisual speech adaptation than following visual speech adaptation.

We employed the /b/-/v/ contrast for our test stimuli. Previous work has determined that the contrast between /b/ and /v/ is especially salient visually (e.g., Rosenblum & Saldaña, 1992, 1996; Saldaña & Rosenblum, 1994). These contrasting syllables have also been previously found to successfully adapt within (Dias et al., 2016; Saldaña & Rosenblum, 1994) and across the senses (at least for /ba/, Dias et al., 2016; Chapter 2).

If selective speech adaptation is influenced by modality-neutral information available across the auditory and visual modalities, then changes in auditory speech identification following audiovisual speech adaptation should exceed changes following auditory speech adaptation. Similarly, changes in visual speech identification following audiovisual speech adaptation should exceed changes following visual speech adaptation.

Method

Participants

For this experiment, 229 undergraduates (121 Female, 108 Male; M_{age} = 19.24 years, SE_{age} = .080) from the University of California, Riverside participated for course credit. All participants were native English speakers with normal hearing and normal (or corrected-to-normal) vision. Participants were randomly assigned to one of eight experimental groups.

Materials

Test Continua. The same test-stimulus continua used by Dias and Rosenblum (2016; Chapter 2) were employed for the current investigation. These test-stimulus continua were constructed from the utterances of two male speakers (both 32-year-old native English speaking; Speaker S1 a California native, Speaker S2 an Indiana native), who were digitally audio-video recorded uttering /ba/ and /va/ (for details, see Dias & Rosenblum, 2016; Chapter 2).

Auditory test-continua. An eleven-item auditory /va/-to-/ba/ continuum of test-stimuli was made for each speaker by first digitally extracting the auditory components of their /va/ utterances. These /va/ utterances were then made into a /va/-to-/ba/ continuum by systematically deleting pieces of the /va/ fricative and replacing it with silence (for details, see Dias & Rosenblum, 2016; Chapter 2). All continuum items were 1.604s in length. Pilot tests found that these test-stimuli were unambiguously identified as /va/ at the end of the continuum with the longest fricative and unambiguously identified as /ba/

at the end of the continuum with the shortest fricative. Test-stimuli in the middle of the continua were identified half of the time as /va/ and half of the time as /ba/.

Visual Test Continua. An eleven-item visual /va/-to-/ba/ continuum of test-stimuli was made for each speaker by first digitally extracting the visual components of the speaker's /va/ and /ba/ utterances. Each speaker's /va/ utterance was then digitally superimposed over their /ba/ utterance. Eleven copies of each stimulus were then created. Each of the eleven copies was then modified to decrease the opacity of the superimposed /va/ utterance in 10% increments. The finished stimuli were each 1.604s in length, corresponding with the auditory test continua (for details, see Dias & Rosenblum, 2016; Chapter 2). Pilot tests found the test stimuli to be unambiguously identified as /va/ at the end of the continuum where opacity of the superimposed /va/ utterance was 100% and unambiguously identified as /ba/ at the end of the continuum where the opacity of the superimposed /va/ utterance was 0%. Test stimuli in the middle of the continua (50% opacity) were identified half of the time as /va/ and half of the time as /ba/.

Adaptors. The same unimodal auditory and visual adaptors used by Dias and Rosenblum (2016; Chapter 2) were employed for the current investigation. The /va/ and /ba/ endpoints of the auditory continuum served as auditory adaptors, and the /va/ and /ba/ endpoints of the visual continuum served as visual adaptors (for details, see Dias & Rosenblum, 2016; Chapter 2).

For the current investigation, we also constructed bimodal audiovisual-/ba/ and audiovisual-/va/ adaptors. To construct these stimuli, we combined the congruent auditory and visual adaptors by dubbing the auditory utterances onto their respective visual stimulus components. We dubbed the auditory and visual components by first lining up the auditory utterance with the visible speech articulation in the visual component. We then made fine-tuned adjustments to ensure that there was no perceptible asynchrony between the auditory and visual components. The finished audiovisual-/ba/ and audiovisual-/va/ stimuli served as our critical test adaptors.

Procedure

The general procedure employed for this study was the same as that used by Dias and Rosenblum (2016; Chapter 2). The 229 participants were randomly assigned to eight groups (approximately 28 participants per group). Four different groups were tested for changes in auditory speech identification (using the auditory test-continuum) following adaptation to either 1) auditory-/ba/, 2) auditory-/va/, 3) audiovisual-/ba/, or 4) audiovisual-/va/. Similarly, four different groups were tested for changes in visual speech identification (using the visual test-continuum) following adaptation to either 1) visual-/ba/, 2) visual-/va/, 3) audiovisual-/ba/, or 4) audiovisual-/va/.

Results

Participant responses were coded as the proportion of times they identified each of the 11 stimuli from the test continuum as /ba/ before and after adaptation (see *Figure 1*). Using the method of probits (Finney, 1971), normal cumulative ogives were fit to each participants identification functions before and after adaptation. The number corresponding with the hypothetical test stimulus at the 50% point of each participant's identification functions served as the perceived phonemic boundary between /ba/ and /va/ along the test-continuum.

To determine whether adaptation significantly changed how stimuli along the test-continua were identified, 2-tailed paired-samples t-tests were employed comparing phonemic boundaries before and after adaptation (e.g., Dias et al., 2016; Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994). *Table 3.2* reports the results. As can be seen, all groups exhibited a significant change in test-stimulus identification following adaptation, regardless of whether the adaptor was unimodal (auditory or visual) or bimodal (audiovisual).

Comparing the size of the change in auditory and visual speech identification following unimodal and bimodal adaptation revealed that the congruent crossmodal information provided in bimodal adaptors did not produce greater adaptation over unimodal adaptors. In fact, the results suggest that adaptation to audiovisual-/va/ produced *smaller* changes in visual speech identification than adaptation to visual-/va/ (see *Table 3.3*). The results suggest that changes in auditory and visual speech identification when adapting and testing in the same modality is not enhanced by simultaneous adaptation to congruent crossmodal speech information.

Discussion

Dias and Rosenblum (2016; Chapter 2) previously suggested that changes in speech perception that follow within-modality and crossmodal adaptation may be based on the same modality neutral articulatory information. However, they found that changes in speech perception are greater following within-modality adaptation, suggesting that there is more information overlap when adapting and testing in the same modality than when adapting and testing in different modalities. These differences between within-modality and crossmodal adaptation may give some clue as to why we failed to observe bimodal enhancement of selective speech adaptation in the current investigation. One possibility is that the changes in speech perception that follow within-modality adaptation have an upper-limit, a point at which any further adaptation will no longer change speech perception. If that limit is reached following within-modality adaptation, any influence of simultaneous crossmodal adaptation would be non-observable. If such a limit can account for our failure to observe a bimodal enhancement of selective speech adaptation, then methodological techniques could be used to control for such effects in the future.

Studies in the domains of speech comprehension (e.g., Erber, 1975; Ross et al., 2007; Sumby & Pollack, 1954), talker identification (e.g., Remez, Fellowes, & Rubin, 1997), and speech imitation (e.g., Dias & Rosenblum, 2015) have controlled for ceiling effects and apparent upper-limits when investigating audiovisual bimodal enhancement by masking or distorting some of the information in the speech signal. For example, the accurate identification of clear auditory and audiovisual speech is typically near perfect. Adding white noise to an auditory speech signal can mask some of the information in the signal, making accurate identification of auditory speech more difficult. However, accurate identification of auditory speech in noise can be improved when the articulating face of the speaker is visible, demonstrating a bimodal enhancement in speech identification (Erber, 1975; Ross et al., 2007; Sumby & Pollack, 1954).

Adding noise to an adaptor stimulus may help to control for a possible upper-limit in selective speech adaptation by masking some of the information shared between an adaptor and a test-stimuli. For example, if noise can mask some of the information shared between adaptors and test-stimuli of the same modality, then changes in auditory speech identification following adaptation to an auditory adaptor in noise will be less than changes following adaptation to a clear auditory adaptor. The smaller changes in auditory speech identification following adaptation to auditory speech in noise might then be improved by adding a visual speech component to the auditory adaptor in noise. Follow-up investigations should explore this possibility.

Alternatively, a ceiling effect might be avoided by using bimodal adaptors that incorporate sensory components *not* shared with test-stimuli. Remember that (Dias & Rosenblum, 2016; Chapter 2) suggested that less information is shared between adaptors and test-stimuli of different modalities than between adaptors and test-stimuli of the same modality. Remember also that Cooper and colleagues have already demonstrated that speech production (and the kinesthetic and tactile sensations associated with it) can change auditory speech identification (Cooper, 1974; Cooper, Billings, & Cole, 1976; Cooper, Blumstein, & Nigro, 1975). It may be that changes in auditory speech identification following bimodal adaptation to visual *and* kinesthetic/tactile (speech production) information may be greater than changes following adaptation to unimodal visual information. Along with the noise manipulation discussed previously, follow-up investigations should also explore bimodal enhancements in selective speech adaptation when bimodal adaptors and test-stimuli have no overlapping sensory components.

References

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92*(2), 339-355.

Baart, M., & Vroomen, J. (2010). Do you see what you are hearing? Cross-modal effects of speech sounds on lipreading. *Neuroscience Letters, 471,* 100-103.

Bernstein, L. E., Auer, E. T. J., Eberhardt, S. P., & Jiang, J. (2013). Auditory perceptual learning for speech perception can be enhanced by auditory training. *Frontiers in Neuroscience, 7*(34). doi:10.3389/fnins.2013.00034

Cooper, W. E. (1974). Perceptuomotor adaptation to a speech feature. *Perception & Psychophysics, 16*(2), 229-234.

Cooper, W. E. (1975). Selective Adaptation to Speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive Theory* (Vol. 1, pp. 23-54). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooper, W. E., Billings, D., & Cole, R. A. (1976). Articulatory effects on speech perception: A second report. *Journal of Phonetics, 4*(3), 219-232.

Cooper, W. E., Blumstein, S. E., & Nigro, G. (1975). Articulatory effects on speech perception: A preliminary report. *Journal of Phonetics, 3,* 87-98.

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. doi:10.1016/j.wocn.2016.02.004

Dias, J. W., & Rosenblum, L. D. (2011). Visual influences on interactive speech alignment. *Perception, 40,* 1457-1466.

Dias, J. W., & Rosenblum, L. D. (2015). Visibility of speech articulation enhances auditory phonetic convergence. *Attention, Perception, & Psychophysics, 77*(6). doi:10.3758/s13414-015-0982-6

Dias, J. W., & Rosenblum, L. D. (2016). Selective adaptation of crossmodal speech information is not the result of higher-level stimulus associations. *The Journal of the Acoustical Society of America, 139*(4), 2017-2018.

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

Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders, 40*(4), 481-492.

Finney, D. J. (1971). *Probit Analysis*. Cambridge, MA: Cambridge University Press.

Fowler, C. A. (2004). Speech as a supramodal or amodal phenomenon. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processing* (pp. 189-202). Cambridge, MA: MIT Press.

Ganong, W. F. (1978). The selective adaptation effects of burst-cued stops. *Perception & Psychophysics, 24*(1), 71-83.

Hardison, D. M. (2005). Variability in bimodal spoken language processing by native and nonnative speakers of English: A closer look at effects of speech style. *Speech Communication, 46*(1), 73-93. doi:http://dx.doi.org/10.1016/j.specom.2005.02.002

Jones, B. C., Feinberg, D. R., Bestelmeyer, P. E. G., DeBruine, L. M., & Little, A. C. (2010). Adaptation to different mouth shapes influences visual perception of ambiguous lip speech. *Psychonomic Bulletin & Review, 17*(4), 522-528.

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

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research, 71*, 4-12.

Paris, T., Kim, J., & Davis, C. (2013). Visual speech form influences the speed of auditory speech processing. *Brain and Langauge, 126*, 350-356.

Remez, R. E. (1980). Susceptibility of a stop consonant to adaptation on a speech-nonspeech continuum: Further evidence against feature detectors in speech perception. *Perception & Psychophysics, 27*(1), 17-23.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 651-666.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics, 30*(4), 309-314.

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science, 17*(6), 405-409.

Rosenblum, L. D., Dorsi, J. J., & Dias, J. W. (2016). The impact and status of Carol Fowler's supramodal theory of multisensory speech perception. *Ecological Psychology*.

Rosenblum, L. D., Miller, R. M., & Sanchez, K. (2007). Lip-read me now, hear me better later: Cross-modal transfer of talker-familiarity effects. *Psychological Science, 18*(5), 392-396.

Rosenblum, L. D., & Saldaña, H. M. (1992). Discrimination tests of visually influenced syllables. *Perception and Psychophysics, 52*(4), 461-473.

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 318-331.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17,* 1147-1153.

Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *Journal of the Acoustical Society of America, 95*(6), 3658-3661.

Samuel, A. G., & Kat, D. (1998). Adaptation is Automatic. *Perception & Psychophysics, 60*(3), 503-510.

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), 1479-1490.

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*(4), 661-699.

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

Vroomen, J., & Baart, M. (2012). Phonetic recalibration in audiovisual speech. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes* (pp. 363-379). Boca Raton, FL: CRC Press.

Table 3.1
Changes in speech categorization following unimodal and bimodal speech adaptation in previous studies

	Test-Continuum	Unimodal Adaptor	Unimodal Shift	Bimodal Adaptor	Bimodal Shift	Difference in Shifts
Roberts & Summerfield (1981)	Auditory /ba/-to-/da/	Auditory-/ba/	2.213	Audiovisual-/ba/	2.030	-0.183
	Auditory /ba/-to-/da/	Auditory-/da/	1.568	Audiovisual-/da/	1.697	0.128
Dias & Rosenblum (2016)	Audiovisual /ba/-to-/va/	Visual-/va/	1.386	Audiovisual-/va/	1.987	0.601

Notes: The unimodal and bimodal shift values represent the absolute shift in phonemic category boundary following adaptation. Positive shift differences between unimodal and bimodal adaptors suggest a bimodal enhancement in selective speech adaptation.

Table 3.2
Phonemic category boundaries and magnitudes of boundary shifts following adaptation.

Continuum		Adaptor	Phonemic Boundary								
Modality	Adaptor	Modality	Baseline	Adapted	Shift	*SE*	*t*	*n*	$p_{(1\text{-}tailed)}$		$r_{effect\ size}$
Auditory	/ba/	AO	5.681	6.839	1.158	0.252	4.598	26	0.000	***	0.677
		AV	5.845	6.882	1.037	0.255	4.072	30	0.000	***	0.603
	/va/	AO	5.816	4.137	-1.679	0.231	-7.255	26	0.000	***	0.823
		AV	5.589	4.095	-1.494	0.317	-4.713	30	0.000	***	0.659
Visual	/ba/	VO	6.042	6.481	0.440	0.093	4.75	29	0.000	***	0.668
		AV	5.771	6.204	0.434	0.111	3.895	30	0.001	***	0.586
	/va/	VO	5.925	5.301	-0.623	0.134	-4.656	28	0.000	***	0.667
		AV	5.929	5.667	-0.261	0.098	-2.67	30	0.006	**	0.444

Notes: *p < .05, ***p < .001. AO = Auditory, VO = Visual, AV = Audiovisual.

Table 3.3

Differences between the phonemic boundary shifts following bimodal and unimodal adaptation.

Continuum			Difference in Phonemic						
Modality	Bimodal Adaptor	Unimodal Adaptor	Boundary Shift	*SE*	*t*	*n*	$p_{(2\text{-tailed})}$		$r_{effect\ size}$
Auditory	Audiovisual-/ba/	Auditory-/ba/	-0.120	0.360	-0.333	56	0.740		-0.045
	Audiovisual-/va/	Auditory-/va/	0.185	0.403	0.460	56	0.648		-0.062
Visual	Audiovisual-/ba/	Visual-/ba/	-0.006	0.145	-0.039	59	0.969		-0.005
	Audiovisual-/va/	Visual-/ba/	0.362	0.164	2.203	58	0.032	*	-0.282

Notes: *$p < .05$. Differences in phonetic boundary shift are calculated from the means reported in Table 3.1. Negative *r*-values represent counter-hypothetical effects.

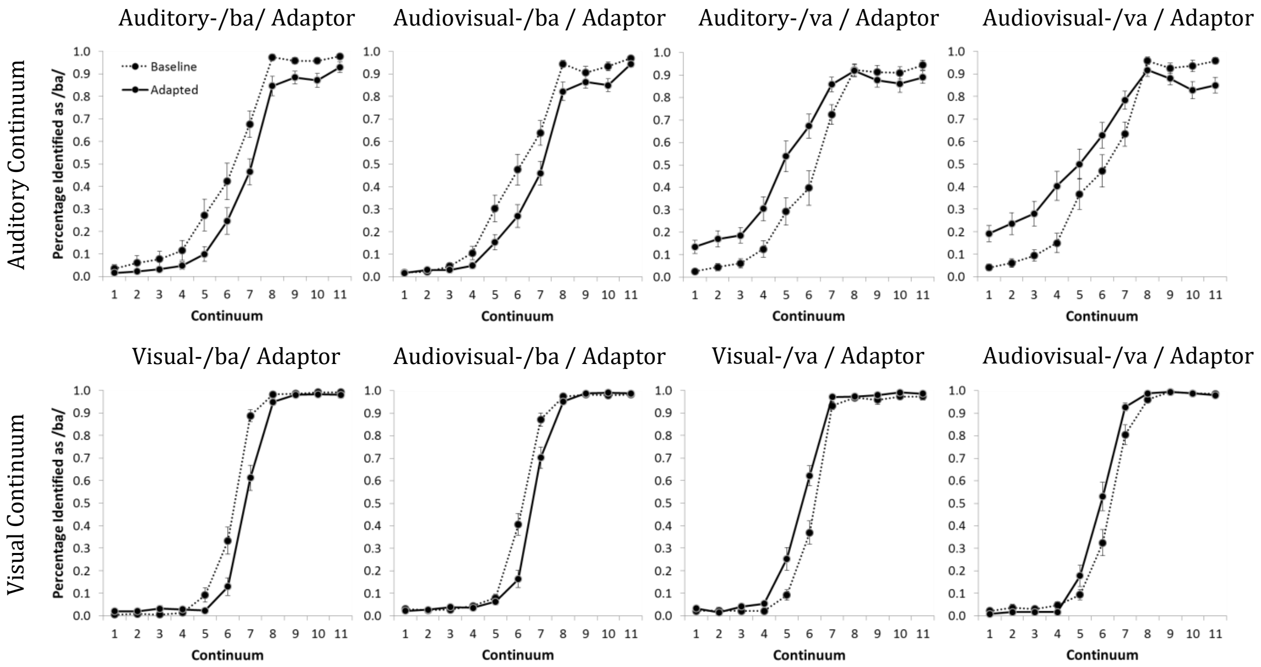


Figure 3.1. The average proportion each continuum test-stimulus was identified as /ba/ before (Baseline, dashed line) and after (Adapted, solid line) adaptation. Data is represented for each of the eight groups, depending on tested continuum modality and adaptor. Error bars represent standard errors of the mean.

Conclusion

Crossmodal Influences in Selective Speech Adaptation

The results of the current series of investigations provide some new insights regarding the underlying basis of selective speech adaptation. Counter to previous accounts suggesting selective speech adaptation depends on sense-specific spectrotemporal overlap between adaptors and test stimuli (e.g., Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994), the results of Chapter 2 suggests that selective speech adaptation is sensitive to modality-neutral information. However, crossmodal adaptation is subtle compared to within-modality adaptation (when adapting and testing in the same sensory modality), requiring larger samples of participants to reach statistical significance.

The subtleness of crossmodal adaptation may account for our failure to find evidence for crossmodal adaptation in Chapter 1. Recall that Chapter 1 investigated the influences of crossmodal adaptation on the identification of visually influenced audiovisual percepts. Crossmodal influences were non-significant. However, the subtleness of crossmodal adaptation observed in Chapter 2, combined with the small sample sizes employed in Chapter 1 (consistent with previous investigations of crossmodal speech adaptation; e.g., Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994), may account for the non-significant crossmodal effects.

Chapter 2 offers competing theories of selective speech adaptation that may account for the subtleness of crossmodal adaptation. Subtle crossmodal adaptation, compared to within-modality adaptation, may suggest that different perceptual mechanisms are responsible for processing sensory-specific (auditory or visual) information and modality-neutral (articulatory) information. It may be that adaption affects both sensory-specific and modality-neutral mechanisms. When testing in the same modality that was adapted, greater adaptation may be observed because both the sensory-specific and modality-neutral mechanisms have been affected. However, when testing in a modality different from the modality adapted, less adaptation may be observed because *only* the modality-neutral mechanism has been affected.

Alternatively, it may be that a single modality-neutral mechanism is responsible for selective speech adaptation. The subtleness of crossmodal adaptation may suggest that adaptors and test stimuli of different modalities share less overlapping information than adaptors and test-stimuli of the same modality. For example, while the place of articulation between /b/ and /v/ can be seen and heard, the vocal tone of these utterances is something heard but not seen (e.g., Summerfield, 1987). Similar explanations have been provided to account for the differences in within-modality and crossmodal talker-familiarity effects in speech comprehension (e.g., Rosenblum, Miller, & Sanchez, 2007).

Either theory of selective speech adaptation may account for the failure to find bimodal enhancement of speech adaptation in Chapter 3. Recall that Chapter 3 investigated whether changes in speech identification were greater following bimodal audiovisual adaptation than following unimodal adaptation (when the adaptor and test-stimuli were of the same sensory modality). Bimodal adaptation failed to produce stronger adaptation effects compared to unimodal adaptation. However, as discussed in Chapter 3, possible upper-limits in selective speech adaptation cannot be discounted.

These limits could result from adaptation simultaneously affecting sensory-specific and modality-neutral mechanisms. Alternatively, it may be that the greater amount of overlapping information between adaptors and test-stimuli of the same modality can account for upper-limits in selective speech adaptation. Methods to control for such limits in future investigations were proposed, including introducing noise to the adaptor stimulus and using bimodal adaptors consisting of sensory components not shared with test-stimuli.

Regardless of the underlying mechanism(s), crossmodal selective speech adaptation seems to depend (at least in part) on the lawful relationship between adaptors and test-stimuli. Remember that adaptation to text in Experiment 3 (Chapter 2), which does not share lawful (articulatory) information with heard and seen speech, failed to produce changes in speech identification consistent with selective speech adaptation. The results of Experment 3 (Chapter 2) also suggest that the changes in speech identification following crossmodal adaptation are not likely the result of some (cognitive) change in response bias (e.g., Storrs, 2015; Storrs & Arnold, 2012) or to the subvocalization of read text.

Implications for Prior Theoretical Accounts

The account provided by Roberts and Summerfield (1981) suggesting that selective speech adaptation depends on sense-specific spectrotemporal overlap between adaptors and test stimuli may not be entirely correct. Though sensory-specific mechanisms may be involved, selective speech adaptation seems to be susceptible to crossmodal influence (Chapter 2).

Similarly, the account provided by Saldaña and Rosenblum (1994) (see also Dias, Cook, & Rosenblum, 2016; Chapter 1) suggesting that selective speech adaptation occurs at a level of processing prior to the (complete) integration of information across sensory modalities may not be correct. From some theoretical perspectives, integration of information across sensory modalities occurs at the information source, such that information is integrated across sensory modalities because the information is lawfully association with a common event (e.g., Gibson, 1966). From this perspective, *any* crossmodal influences in perceptual processing of speech may indicate that information available across modalities is integrated (e.g., Fowler, 2004; Rosenblum, 2005; Rosenblum, Dorsi, & Dias, 2016). Finding crossmodal adaptation in Chapter 2 suggests that information available across the auditory and visual modalities integrates prior to the level of selective speech adaptation. As such, the results of Chapter 2 are consistent with evidence from other speech domains suggesting crossmodal influences occur early the speech process (for a review, see Rosenblum et al., 2016).

The role of top-down influences in selective speech adaptation (Samuel, 1997, 2001; Samuel & Lieblich, 2014) requires consideration in the face or crossmodal adaptation. The initial premise of the theoretical account offered by Samuel and Lieblich (2014) may be correct: Selective speech adaptation is primarily affected by sense-specific bottom-up information. The subsequent premise of the account offered by Samuel and Lieblich (2014) may also be correct, that top-down knowledge can affect selective speech adaptation in the absence of bottom-up information. However, if crossmodal influences in selective speech adaptation are to be considered as top-down influences (i.e., Samuel and Lieblich, 2014), then the distinction that Samuel & Lieblich (2014) make between visual and lexical information (visual as perceptual; lexical as perceptual *and* linguistic) may be incorrect, since *both* visual (crossmodal) and lexical information can affect selective speech adaptation. It should be noted, however, that the notion of crossmodal visual adaptation as a top-down influence is not supported by the broader literature on multisensory speech processing (e.g., Rosenblum, et al., 2016), or the results of text adaptation (Chapter 2, Experiment 3).

Alternatively, Samuel and Lieblich's (2014) conceptualization of bottom-up information may be incorrect. Though they suggest that bottom-up information is sense-specific, the results of Chapter 2 may suggest that bottom-up information can be modality-neutral (e.g., Gibson, 1966). The subsequent premise of the account offered by Samuel and Lieblich (2014), that top-down knowledge can affect selective speech adaptation in the absence of bottom-up information, would still be consistent with the influences of lexical knowledge on selective speech adaptation (Samuel, 1997, 2001; Samuel & Lieblich, 2014).

In general, this dissertation (Chapter 2 in particular) provides some evidence suggesting that selective speech adaptation can be subtly influenced by information available across the auditory and visual modalities.

References

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. doi:10.1016/j.wocn.2016.02.004

Fowler, C. A. (2004). Speech as a supramodal or amodal phenomenon. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processing* (pp. 189-202). Cambridge, MA: MIT Press.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin Company.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics*, *30*(4), 309-314.

Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. Pisoni & R. Remez (Eds.), *Handbook of Speech Perception* (pp. 51-78). Malden: Blackwell.

Rosenblum, L. D., Dorsi, J. J., & Dias, J. W. (2016). The impact and status of Carol Fowler's supramodal theory of multisensory speech perception. *Ecological Psychology*.

Rosenblum, L. D., Miller, R. M., & Sanchez, K. (2007). Lip-read me now, hear me better later: Cross-modal transfer of talker-familiarity effects. *Psychological Science, 18*(5), 392-396.

Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *Journal of the Acoustical Society of America, 95*(6), 3658-3661.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32*, 97-127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*(4), 348-351.

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), 1479-1490.

Storrs, K. R. (2015). Are high-level aftereffects perceptual? *Frontiers in Psychology, 6*, 157. doi:10.3389/fpsyg.2015.00157

Storrs, K. R., & Arnold, D. H. (2012). Not all face aftereffects are equal. *Vision Research, 64*, 7-16. doi:http://dx.doi.org/10.1016/j.visres.2012.04.020