Title

Influences of selective adaptation on perception of audiovisual speech

Authors

Dias, James W
Cook, Theresa C
Rosenblum, Lawrence D

Publication Date

2016-05-01

DOI

10.1016/j.wocn.2016.02.004

Peer reviewed

HHS Public Access

Author manuscript

J Phon. Author manuscript; available in PMC 2017 May 01.

Published in final edited form as:

J Phon. 2016 May ; 56: 75–84. doi:10.1016/j.wocn.2016.02.004.

Influences of selective adaptation on perception of audiovisual speech

James W. Dias[*], **Theresa C. Cook**, and **Lawrence D. Rosenblum**

University of California, Riverside

Abstract

Research suggests that selective adaptation in speech is a low-level process dependent on sensory-specific information shared between the adaptor and test-stimuli. However, previous research has only examined how adaptors shift perception of unimodal test stimuli, either auditory or visual. In the current series of experiments, we investigated whether adaptation to cross-sensory phonetic information can influence perception of integrated audio-visual phonetic information. We examined how selective adaptation to audio and visual adaptors shift perception of speech along an *audiovisual* test continuum. This test-continuum consisted of nine audio-/ba/-visual-/va/ stimuli, ranging in visual clarity of the mouth. When the mouth was clearly visible, perceivers "heard" the audio-visual stimulus as an integrated "va" percept 93.7% of the time (e.g., McGurk & MacDonald, 1976). As visibility of the mouth became less clear across the nine-item continuum, the audio-visual "va" percept weakened, resulting in a continuum ranging in audio-visual percepts from /va/ to /ba/. Perception of the test-stimuli was tested before and after adaptation. Changes in audiovisual speech perception were observed following adaptation to visual-/va/ and audiovisual-/va/, but not following adaptation to auditory-/va/, auditory-/ba/, or visual-/ba/. Adaptation modulates perception of integrated audio-visual speech by modulating the processing of sensory-specific information. The results suggest that auditory and visual speech information are not completely integrated at the level of selective adaptation.

Keywords

McGurk effect; speech perception; selective adaptation; crossmodal; cross-sensory; sensory; integration

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,

[*]Correspondences should be addressed to James W. Dias, University of California, Department of Psychology, 900 University Ave., Riverside, CA 92521.

& 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 visual Bernstein, Auer, & Moore, 2004; 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 auditoryphonetic perception following adaptation to audio-visual 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 audio-visual 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, Magnotti, & Beauchamp, 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 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 test-stimuli 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 audiovisual 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 cross-sensory 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 visual-speech 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.

However, if selective adaptation modulates perception of audio-visual speech by influencing processing of shared sensory-specific phonetic information between an adaptor and the test stimuli, then visual-/va/ and audiovisual-/va/ should be the *only* adaptors to produce perceptual shifts. Visual-/va/ and audiovisual-/va/ are the only adaptors tested that share sensoryspecific phonetic information expected to shift perception of ambiguous phonetic information in the audiovisual test continuum (see Table 1). These results would suggest that auditory and visual speech information do not fully integrate by the time information reaches the level of selective adaptation.

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 640×480 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 Figure 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 1,800ms 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 1,800ms 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 (1,100ms). 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 640×480 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 640×480 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 (24in ViewSonic VX2450 at 60Hz and 1920×1080 resolution) and headphones (Sony MDR-V600 headphones adjusted to 70dB 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 every three trials one A-/ba/-V-/va/, one AV-/ba/, and one AV-/va/ stimulus was presented. Each stimulus was presented 5 times over the course of 135 trials ([9 (A-/ba/-V-/va/) + 9 (AV-/ba/) + 9 (AV-/va/)] × 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 (100ms 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 Figure 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 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 ($p < .05$) 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 & Summerfield (1981) and Saldaña & 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, & Rosenblum, in press; 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, & Vatikiotis-Bateson, 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, Sams, Nicol, & Kraus, 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.

Acknowledgments

This research was supported by NIDCD Grant 1R01DC008957-01.

References

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

Baart M, Vroomen J. Do you see what you are hearing? Cross-modal effects of speech sounds on lipreading. Neuroscience Letters. 2010; 471:100–103. [PubMed: 20080146]

Bernstein, LE.; Auer, ETJ.; Moore, JK. Audiovisual speech binding: Convergence or association. In: Calvert, GA., editor. The Handbook of Multisensory Processes. Cambridge, MA: MIT Press; 2004. p. 203-223.

Callan DE, Callan AM, Kroos C, Vatikiotis-Bateson E. Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. Cognitive Brain Research. 2001; 10:349–353. [PubMed: 11167060]

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, David AS. Activation of auditory cortex during silent lipreading. Science. 1997; 276:593–596. [PubMed: 9110978]

Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Current Biology. 2000; 10(11):649–657. [PubMed: 10837246]

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

Colin C, Radeau M, Soquet A, Demolin D, Colin F, Deltenre P. Mismatch negativity evoked by the McGurk-MacDonald effect: A phonetic representation within short-term memory. Clinical Neurophysiology. 2002; 113:495–506. [PubMed: 11955994]

Dias, JW.; Cook, TC.; Rosenblum, LD. The McGurk effect and the primacy of multisensory perception. In: Shapiro, AG.; Todorovic, D., editors. Oxford Compendium of Visual Illusions. Oxford University Press; (in press)

Eimas PD, Cooper WE, Corbit JD. Some properties of linguistic feature detectors. Perception & Psychophysics. 1973; 13(2):247–252.

Eimas PD, Corbit JD. Selective adaptation of linguistic feature detectors. Cognitive Psychology. 1973; 4:99–109.

Erber NP. Auditory-visual perception of speech. Journal of Speech and Hearing Disorders. 1975; 40(4):481–492. [PubMed: 1234963]

Finney, DJ. Probit Analysis. Cambridge, MA: Cambridge University Press; 1971.

Fowler, CA. Speech as a supramodal or amodal phenomenon. In: Calvert, GA.; Spence, C.; Stein, BE., editors. The handbook of multisensory processing. Cambridge, MA: MIT Press; 2004. p. 189-202.

Fowler CA, Dekle DJ. Listening with eye and hand: Cross-modal contributions to speech perception. Journal of Experimental Psychology: Human Perception and Performance. 1991; 17(3):816–828. [PubMed: 1834793]

Ganong WF. The selective adaptation effects of burst-cued stops. Perception & Psychophysics. 1978; 24(1):71–83. [PubMed: 693243]

Gick B, Derrick D. Aero-tactile integration in speech perception. Nature. 2009; 426:502–504. [PubMed: 19940925]

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychological Review. 1998; 105(2):251–279. [PubMed: 9577239]

Green, KP. The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In: Campbell, R.; Dodd, B., editors. Hearing by Eye II: Advances in the Psychology of Speechreading and Audiovisual Speech. Hove, UK: Psychology Press; 1998. p. 3-25.

Green KP, Miller JL. On the role of visual rate information in phonetic perception. Perception & Psychophysics. 1985; 38(3):269–276. [PubMed: 4088819]

Green KP, Norrix LW. Perception of /r/ and /l/ in a stop cluster: Evidence of cross-modal context effects. Journal of Experimental Psychology: Human Perception and Performance. 2001; 27(1): 166–177. [PubMed: 11248931]

Ito T, Tiede M, Ostry DJ. Somatosensory function in speech perception. PNAS. 2009; 106(4):1245–1248. [PubMed: 19164569]

Jones BC, Feinberg DR, Bestelmeyer PEG, DeBruine LM, Little AC. Adaptation to different mouth shapes influences visual perception of ambiguous lip speech. Psychonomic Bulletin & Review. 2010; 17(4):522–528. [PubMed: 20702872]

Kat D, Samuel AG. More adaptation of speech by nonspeech. Journal of Experimental Psychology: Human Perception and Performance. 1984; 10(4):512–525. [PubMed: 6235316]

MacDonald J, McGurk H. Visual influences on speech perception processes. Perception & Psychophysics. 1978; 24(3):253–257. [PubMed: 704285]

Mallick DB, Magnotti JF, Beauchamp MS. Variability and stability in the McGurk effect: contributions of participants, stimuli, time, and response type. Psychonomic Bulletin & Review. 2015:1–9. [PubMed: 24847901]

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

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264:746–748. [PubMed: 1012311]

Miller RM, Sanchez K, Rosenblum LD. Alignment to visual speech information. Attention, Perception, & Psychophysics. 2010; 72(6):1614–1625.

Musacchia G, Sams M, Nicol T, Kraus N. Seeing speech affects acoustic information processing in the human brainstem. Experimental Brain Research. 2006; 168(1–2):1–10. [PubMed: 16217645]

Pardo JS. On phonetic convergence during conversational interaction. Journal of the Acoustical Society of America. 2006; 119(4):2382–2393. [PubMed: 16642851]

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

Reisberg, D.; McLean, J.; Goldfield, A. Easy to hear but hard to understand: A lipreading advantage with intact auditory stimuli. In: Dodd, B.; Campbell, R., editors. Hearing by Eye: The Psychology of Lip-Reading. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.; 1987. p. 97-113.

Remez, RE. Perceptual Organization of Speech. In: Pisoni, DB.; Remez, RE., editors. Handbook of Speech Perception. Oxford: Blackwell; 2005.

Remez RE, Fellowes JM, Pisoni DB, Goh WD, Rubin PE. Multimodal perceptual organization of speech: Evidence from tone analogs of spoken utterances. Speech Communication. 1998; 26:65–73. [PubMed: 21423823]

Roberts M, Summerfield Q. Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. Perception & Psychophysics. 1981; 30(4):309–314. [PubMed: 7322807]

Rosenblum, LD. Primacy of multimodal speech perception. In: Pisoni, D.; Remez, R., editors. Handbook of Speech Perception. Malden: Blackwell; 2005. p. 51-78.

Rosenblum LD. Speech perception as a multimodal phenomenon. Current Directions in Psychological Science. 2008; 17(6):405–409. [PubMed: 23914077]

Rosenblum LD, Saldaña HM. Discrimination tests of visually influenced syllables. Perception and Psychophysics. 1992; 52(4):461–473. [PubMed: 1437479]

Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. Cerebral Cortex. 2007; 17:1147–1153. [PubMed: 16785256]

Saldaña HM, Rosenblum LD. Selective adaptation in speech perception using a compelling audiovisual adaptor. Journal of the Acoustical Society of America. 1994; 95(6):3658–3661. [PubMed: 8046153]

Sams M, Aulanko R, Hamalainen M, Hari R, Lounasmaa OV, Lu S, Simola J. Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. Neuroscience Letters. 1991; 127:141–145. [PubMed: 1881611]

Sams M, Mottonen R, Sihvonen T. Seeing and hearing others and oneself talk. Cognitive Brain Research. 2005; 23:429–435. [PubMed: 15820649]

Samuel AG. Lexical activation produces potent phonemic percepts. Cognitive Psychology. 1997; 32:97–127. [PubMed: 9095679]

Samuel AG. Knowing a word affects the fundamental perception of the sounds within it. Psychological Science. 2001; 12(4):348–351. [PubMed: 11476105]

Samuel AG, Lieblich J. Visual speech acts differently than lexical context in supporting speech perception. Journal of Experimental Psychology: Human Perception and Performance. 2014; 40(4):1479–1490. [PubMed: 24749935]

Sumby WH, Pollack I. Visual contribution of speech intelligibility in noise. Journal of the Acoustical Society of America. 1954; 26:212–215.

Summerfield, Q. Some preliminaries to a comprehensive account of audio-visual speech perception. In: Barbara, I.; Ruth, C., editors. Hearing by eye: The psychology of lipreading. Hillsdale, NJ: Erlbaum; 1987. p. 3-51.

Thomas SM, Jordan TR. Determining the influence of Gaussian blurring on inversion effects with talking faces. Attention, Perception, & Psychophysics. 2002; 64(6):932–944.

Vatikiotis-Bateson E, Eigsti I, Yano S, Munhall KG. Eye movement of perceivers during audiovisual speech perception. Perception & Psychophysics. 1998; 60(6):926–940. [PubMed: 9718953]

Vroomen, J.; Baart, M. Phonetic recalibration in audiovisual speech. In: Murray, MM.; Wallace, MT., editors. The Neural Bases of Multisensory Processes. Boca Raton, FL: CRC Press; 2012. p. 363-379.

Highlights

- Tests of unimodal speech adaptation on perception of integrated audiovisual speech percepts
- Audiovisual speech perception changes follows adaptation to sensory-specific information only
- Auditory and visual speech are not completely integrated at the level of selective adaptation

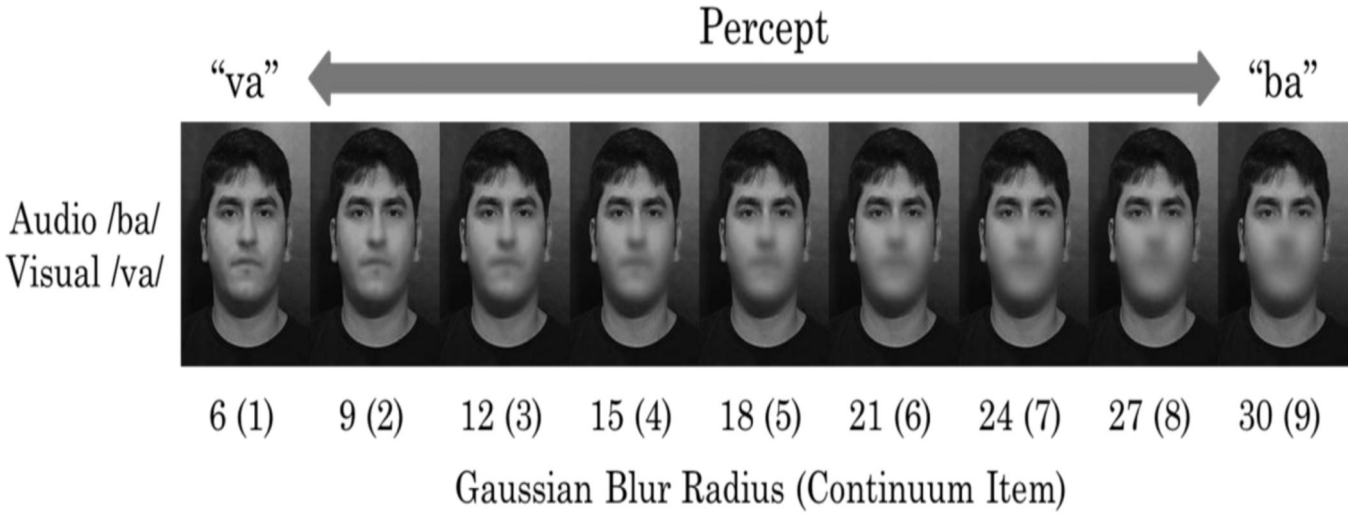

Figure 1.

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.

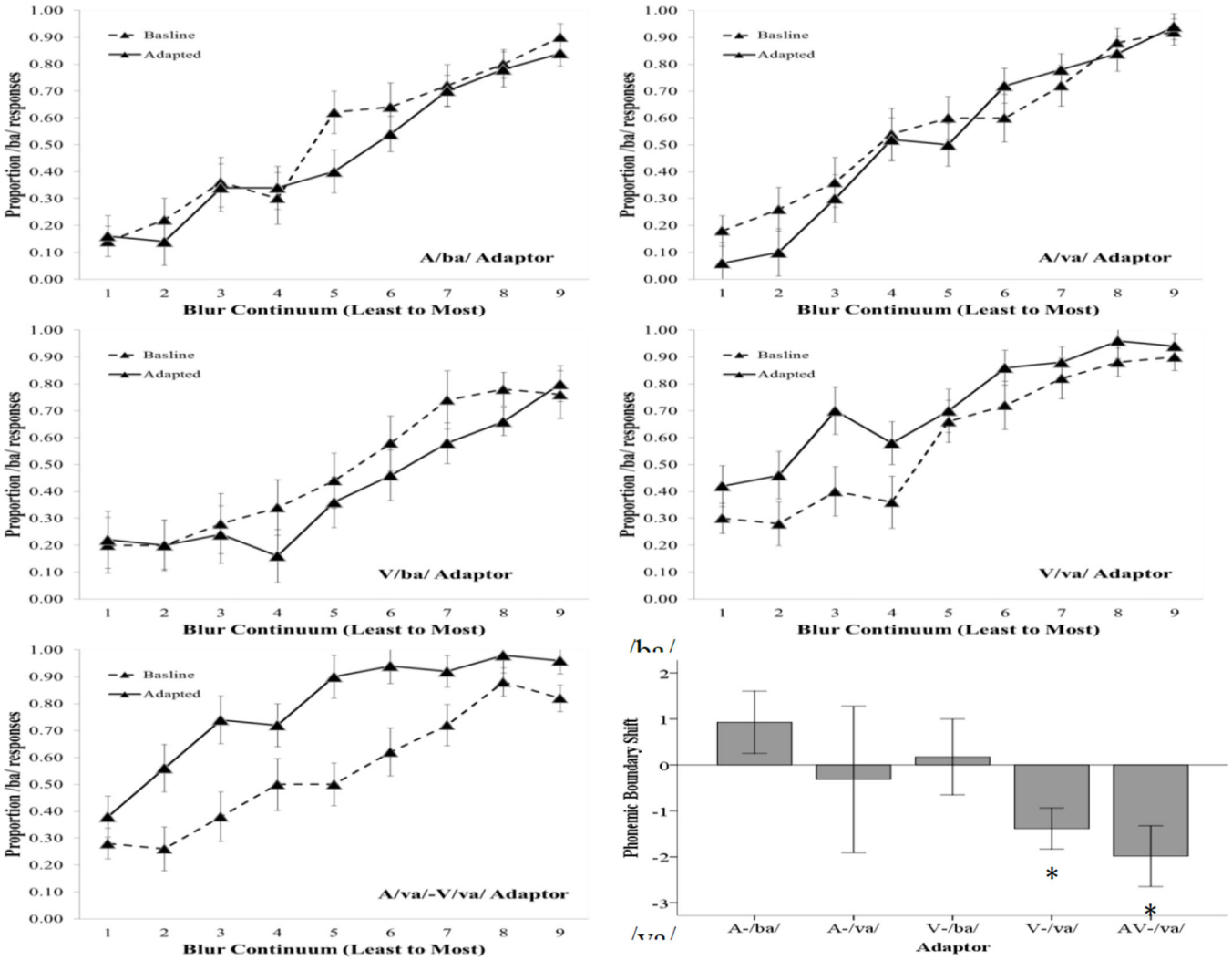

Figure 2.

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.

Table 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 audiovisual speech by affecting processing of sensory specific information shared between the adaptor and audio-visual test-stimuli.

Table 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.