Cognition 147 (2016) 100–105

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Short Communication

Language familiarity modulates relative attention to the eyes and mouth of a talker

Elan Barenholtz [a,*], Lauren Mavica [b], David J. Lewkowicz [c]

[a] Department of Psychology/Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL, United States
[b] Department of Psychology, Florida Atlantic University, Boca Raton, FL, United States
[c] Department of Communication Sciences and Disorders, Northeastern University, Boston, MA, United States

ARTICLE INFO

Article history:
Received 6 March 2015
Revised 8 October 2015
Accepted 23 November 2015
Available online 30 November 2015

Keywords:
Multisensory
Speech
Language
Perception
Attention

ABSTRACT

We investigated whether the audiovisual speech cues available in a talker's mouth elicit greater attention when adults have to process speech in an unfamiliar language vs. a familiar language. Participants performed a speech-encoding task while watching and listening to videos of a talker in a familiar language (English) or an unfamiliar language (Spanish or Icelandic). Attention to the mouth increased in monolingual subjects in response to an unfamiliar language condition but did not in bilingual subjects when the task required speech processing. In the absence of an explicit speech-processing task, subjects attended equally to the eyes and mouth in response to both familiar and unfamiliar languages. Overall, these results demonstrate that language familiarity modulates selective attention to the redundant audiovisual speech cues in a talker's mouth in adults. When our findings are considered together with similar findings from infants, they suggest that this attentional strategy emerges very early in life.

1. Introduction

Speech processing depends on the rapid encoding and interpretation of a complex audiovisual signal. Fortunately, natural languages contain a high degree of structure at the phonetic, lexical, syntactic, and semantic levels and prior knowledge of these structures can facilitate processing. For example, under noisy conditions, perception of auditory speech is more accurate when the spoken language is familiar (Cutler, Weber, Smits, & Cooper, 2004; Gat & Keith, 1978; Lecumberri & Cooke, 2006; Mayo, Florentine, & Buus, 1997; Van Wijngaarden, Steeneken, & Houtgast, 2002). This suggests that language familiarity can reduce the amount of bottom-up information needed to successfully process auditory speech. Here, we asked whether language familiarity also affects responsiveness to audiovisual speech.

Typically, linguistic communication is multisensory in nature. People can both hear and see their interlocutor produce visual and auditory speech signals and they automatically integrate them (McGurk & MacDonald, 1976). Such integration produces a perceptually more salient signal (Meredith & Stein, 1986; Partan & Marler, 1999; Rowe, 1999). Indeed, studies show that concurrent access to redundant audible and visible speech cues enhances speech perception under noisy conditions (Middelweerd & Plomp, 1987; Rosenblum, 2008; Rosenblum, Johnson, & Saldana, 1996; Sumby & Pollack, 1954; Summerfield, 1979). Several recent studies have found that familiarity with a language modulates the perceived timecourse of audiovisual speech: when a language is familiar, the visual speech signal must lead the auditory speech signal by a larger time interval for simultaneity to be perceived compared with when the language is unfamiliar (Love, Pollick, & Petrini, 2012; Navarra, Alsius, Velasco, Soto-Faraco, & Spence, 2010), perhaps because familiarity speeds up the auditory processing of speech.

Language familiarity may also modulate visual selective attention during speech encoding, a possibility supported by evidence from infant studies. Lewkowicz and Hansen-Tift (2012) presented monolingual, English-learning infants of different ages with videos of talkers speaking either in their native language or in a non-native language (i.e., Spanish). At 4 months, infants fixated the talker's eyes, whereas at 8 and 10 months of age—when infants enter the canonical babbling stage and begin to acquire spoken language—they fixated the talker's mouth. At 12 months of age, the infants no longer fixated the mouth more than the eyes when the talker spoke in the infants' native language but continued to fixate the mouth more when the talker spoke in a non-native language.

Lewkowicz and Hansen-Tift's (2012) findings indicated for the first time that selective attention to the audiovisual redundancy

* Corresponding author at: Department of Psychology, Florida Atlantic University, Boca Raton, FL 34331, United States.
E-mail address: elan.barenholtz@fau.edu (E. Barenholtz).

http://dx.doi.org/10.1016/j.cognition.2015.11.013
0010-0277/

available in the mouth is modulated by language familiarity. If this early lip-reading behavior reflects a general encoding strategy in response to language familiarity, these differences in fixation behavior may persist into adult. Of course, as Lewkowicz and Hansen-Tift (2012) noted, lipreading in infancy may reflect acquisition of speech production capacity. If so, the selective deployment of attention to a talker's mouth in infancy may reflect infants' attempt to imitate and produce human speech sounds and, thus, may not generalize to adults. Indeed, Lewkowicz and Hansen-Tift (2012) found in a separate experiment with monolingual English-speaking adults that they looked longer at the eyes of a talker regardless of whether she spoke in their native language or not.

Crucially, the adults in the Lewkowicz and Hansen-Tift (2012) study were only asked to passively watch and listen to the talker. Studies with adults have found, however, that the distribution of attention to the eyes and mouth is modulated by task. For example, findings show that the mouth attracts more attention when speech cues become relevant (Buchan, Paré, & Munhall, 2007; Driver & Baylis, 1993; Lansing & McConkie, 1999, 2003) and especially when the auditory signal is degraded (Driver & Baylis, 1993; Lansing & McConkie, 2003; Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998). Conversely, when the task is to attend to social-reference, emotional, and deictic cues, the eyes attract more attention (Birmingham, Bischof, & Kingstone, 2008; Emery, 2000).

Given these findings, we asked whether speech in an unfamiliar language might cause adults to attend more to a talker's mouth if their explicit task is to process the speech. To test this possibility, we tracked selective attention in adults while they watched and listened to people speaking either in their native and, thus, familiar (English) language or in an unfamiliar (Icelandic or Spanish) language. The participants were explicitly required to encode the speech stimulus by subsequently being asked to perform a simple match-to-sample task. We expected that the participants would attend more to the mouth in the unfamiliar than in the familiar language condition.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Participants were 60 self-described English-speaking monolingual, Florida Atlantic University undergraduate students, participating for course credit. Separate groups of 30 participants, each, were randomly assigned to one of two Language groups (English/Icelandic or English/Spanish). Each group of 30 was further subdivided into two groups of 15 with the order of language presentation (i.e. familiar or unfamiliar first) counterbalanced across participants.

2.1.2. Stimuli

Stimuli consisted of movie recordings of two female models, recorded in a sound-attenuated room and presented on an infrared-based eye tracking system (T60; Tobii Technology, Stockholm, Sweden[1]) on a 17-in. computer monitor. Both models were fully bilingual speakers of both English (with no discernible accent) and one other native language (one Spanish, one Icelandic). Each model was recorded speaking a set of 20 sentences in English and the same 20 sentences in her other, native language. The models were recorded from their shoulders up and were instructed to speak naturally in an emotionally passive tone without moving their head. The face of the models measured approximately 6° visual angle width (ear to ear) by approximately 11° visual angle length. The recorded individual sentences averaged 2.5 s each for all three recorded languages.

2.1.3. Procedure

A single trial is schematized in Fig. 1. Participants were presented with sequentially presented pairs of video segments, each consisting of the same person audibly uttering a short sentence, followed by an audio-only clip of one of the two sentences. Participants had to choose which of the two previously presented audiovisual movie segments corresponded to the audio-only clip. For half the participants, the video sequences consisted of a bilingual female speaking English (familiar) sentences in one block and the same model speaking Icelandic (unfamiliar) sentences in a different block (English/Icelandic group). For the other half of the participants, the sequences consisted of a different model speaking English sentences in one block and the same model speaking Spanish sentences in a different block (English/Spanish group). Participants indicated whether the auditory-only clip was extracted from the first or second movie by pressing a key on the keyboard.

Each participant completed two experimental blocks, each consisting of ten pairs of sentences. In one block, all of the sentences were in English while in the other block they were all in an unfamiliar language, either Icelandic or Spanish. Each group was only presented with one model, speaking both English and Icelandic (Icelandic Group) or English and Spanish (Spanish Group). This ensured that the same visual features were present across the familiar and unfamiliar blocks for each participant. Block order (i.e. familiar or unfamiliar presented first or second) was counterbalanced across participants.

Participants' eye movements were recorded with an eye tracking system (T60; Tobii Technology, Stockholm, Sweden) and analyzed with the Tobii Studio 3.0.6 software. Gaze was monitored using near infrared and both bright and dark pupil-centered corneal reflection. Stimuli were presented on a 17-in. flat panel monitor with a screen resolution of 1280×1024 pixels and a sampling rate of 60 Hz. All participants were tested in a quiet room that was illuminated by the stimulus display and were seated ~60 cm from the screen. A standardized five-point calibration was performed prior to tracking as implemented in Tobii Studio software.

2.1.4. Fixation analyses

We defined three principal areas of interest (AOIs): the mouth, the eyes, and the whole face. For each condition, we calculated the time spent fixating the eye and mouth AOIs as a percentage of the total time spent fixating anywhere within the face AOI (Note that fixations within either the mouth or eyes AOI were counted toward the total fixation duration to the face). Fixation (as contrasted with saccades or other eye movements) durations were determined using Tobii Studio's fixation filter algorithm,[2] which distinguishes between time spent fixating within an AOI (which were the basis of our analyses) and time spent engaging in a saccade (which were not included in the analyses).

2.2. Results

Performance in the matching task was near ceiling (between 95% and 97%) across all conditions and language groups, with no significant difference between familiar and unfamiliar (all p-values >.1 by t-test). Fig. 2A shows the proportion of time spent

[1] Technical specifications are available at: http://www.tobii.com/Global/Analysis/Downloads/User_Manuals_and_Guides/Tobii_T60_T120_EyeTracker_UserManual.pdf.

[2] http://www.tobii.com/eye-tracking-research/global/library/white-papers/the-tobii-i-vt-fixation-filter/.

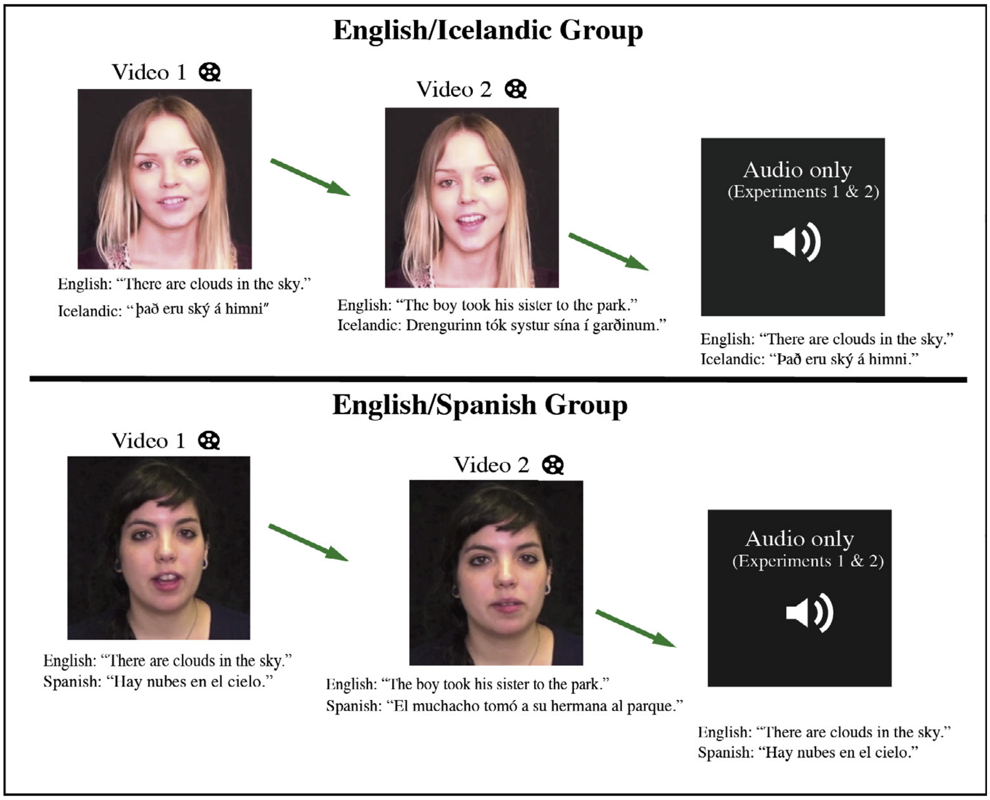

Fig. 1. Schematic representation of the experimental procedure in all three experiments. On each trial, participants were first presented sequentially with two movies showing a person uttering two sentences in either a familiar (English) or unfamiliar (Icelandic or Spanish) language. In Experiments 1 and 2, these movies were followed by an auditory only sample of one of the two sentences previously presented. Participants' task was to report whether the audio-only sentence corresponded to the first or second movie. Experiment 3 did not have an auditory-only experimental task. See text for details.

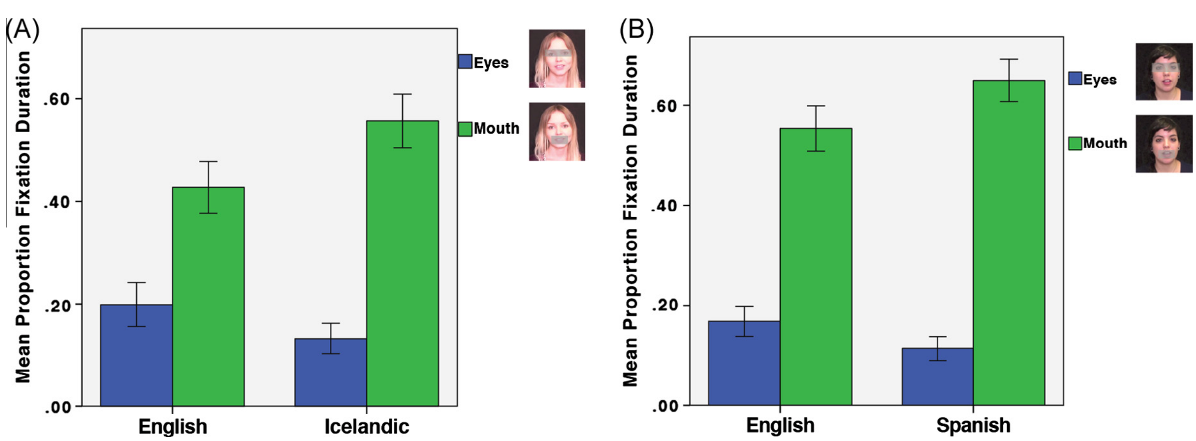

Fig. 2. Proportion of fixation duration (relative to the whole face), for the eyes and mouth AOI's, across languages in the English–Icelandic (A) and English–Spanish (B) blocks in experiment 1. AOIs are shown as transparent gray bars in the face images for illustration; they did not appear in the experimental stimuli. Error bars represent ±1 standard error of the mean.

fixating the mouth and eye regions for the English/Icelandic group. Fig. 2B shows the same results for the English/Spanish group. Consistent with previous studies of selective attention in adults during active speech processing (Driver & Baylis, 1993; Lansing & McConkie, 1999, 2003), we found greater overall fixation of the mouth than the eyes, $t(59) = 12.175$, $p < .001$, $d = 2.35$. In addition, and of particular interest given our initial hypothesis, we found that participants' looked more at the mouth vs. the eyes when the speech was unfamiliar, compared with when it was familiar, in each respective group. Specifically, a paired-samples t-test found that in the English/Icelandic group, the mouth vs. eyes difference score was greater for the Icelandic block of trials, $M = .42$, $SD = .24$ than for the English block of trials, $M = .23$, $SD = .27$, $t(29) = 5.877$, $p < .001$, two-tailed, $d = .74$. Similarly, in the English/Spanish group, the mouth vs. eyes difference score was greater for the Spanish block of trials, $M = .53$, $SD = .26$ than for the English block of trials, $M = .40$, $SD = .30$, $t(29) = 4.596$, $p < .001$, two-tailed, $d = .46$, Thus, across both language pairs, participants looked more at the mouth and less at the eyes when exposed to an unfamiliar language.

3. Experiment 2

Experiment 1 was designed to compare fixation behavior during encoding of familiar and unfamiliar languages. Because we only included monolingual English speakers as participants, English served as the familiar language in both groups. This raises the possibility that some property of the English stimuli, other than familiarity itself, contributed to lower amounts of attention directed at the speaker's mouth compared with the Spanish and Icelandic stimuli. To test whether familiarity *per se* modulates mouth fixations, in Experiment 2 we employed the same task and stimuli as in the English/Spanish group in Experiment 1 except that this time we tested bilingual English/Spanish participants. If language familiarity mediated the increase in attention to the mouth in Experiment 1 – and not some inherent property of the English language stimuli – then bilingual participants, who are equally familiar with both languages, should exhibit equal amounts of attention to the talker's mouth for both languages.

3.1. Methods

3.1.1. Participants

Participants in Experiment 2 were 30[3] self-described Spanish/English bilinguals who reported being equally familiar with both languages.

3.1.2. Stimuli and procedure

The stimuli presented here were only the English and Spanish sentences presented in Experiment 1 and the procedure was identical to the procedure employed in Experiment 1 too.

3.1.3. Results

Fig. 3A shows the proportion of time spent fixating the mouth and eye regions. Analysis of the mouth-vs.-eyes difference scores found no significant difference between the Spanish block of trials, $M = .04$, $SD = .34$, and the English block of trials, $M = .12$, $SD = .33$, $t(29) = 1.705$, $p = .09$, two-tailed. Presumably, these findings reflect the fact that both languages were equally familiar to these participants and, thus, that the encoding task was of equal difficulty. They also indicate that the lower amount of looking at the mouth obtained in response to the English sentences in Experiment 1 was not due to some visual or auditory properties of English *per se* but, rather, to the familiarity of this language to the monolingual, English-speaking participants in that experiment.

Because the English/Spanish stimuli were identical in Experiments 1 and 2, we could compare the difference scores between the monolingual and bilingual participants in the two respective experiments. A 2-way mixed ANOVA on the difference scores, with language as the within-subjects variable and language-expertise as the between-subjects variable, found a significant effect of language (English vs. Spanish), $F(1) = 21.866$, $p < .001$, a significant effect of language-expertise, $F(1) = 18.484$, $p < .001$, and a significant interaction between these two factors, $F(1) = 21.866$, $p < .001$. This interaction supports the hypothesis that familiarity was the determining factor in producing the difference in fixations for the English and Spanish stimuli in Experiment 1.

4. Experiment 3

As indicated earlier, the adults in the Lewkowicz and Hansen-Tift (2012) study looked more at the eyes regardless of language familiarity. As suggested earlier, this may have been due to the fact that their task was to watch and listen to the videos passively. Given our active-processing hypothesis, it is not surprising that, in the absence of an explicit information-processing task, these participants did not deploy greater attentional resources to a talker's mouth. Thus, in this experiment we explicitly investigated whether active processing mediates the greater attention to the mouth found in Experiment 1. To do so, we repeated the English/Icelandic condition from Experiment 1 with monolingual English-speaking adults except that this time participants did not have to perform a specific speech processing task.

4.1. Methods

4.1.1. Participants

We tested a new cohort of 30[4] self-described English-speaking monolinguals.

4.1.2. Stimuli and procedure

Stimuli and procedure were identical to those employed in the English–Icelandic group in Experiment 1. The only difference was that the two movies were not followed by an auditory-only test trial and that the participants were not given any specific task except to freely watch and listen to the movies.

4.1.3. Results and discussion

Fig. 3B shows the proportion of time spent fixating the mouth and eye regions. Analysis of the mouth vs. eyes difference scores found no significant difference between the Icelandic block of trials, $M = .27$, $SD = .41$ and the English block of trials, $M = .22$, $SD = .33$, $p = .174$, two-tailed. Thus, as predicted, when participants were not required to process the audiovisual speech, they did not devote proportionally more selective attention to the mouth in response to the unfamiliar vs. familiar language.

It is interesting to note that we did not obtain greater looking at the eyes in either language condition whereas Lewkowicz and Hansen-Tift (2012) did. This difference may be due to the nature of the stimuli. Here, the stimuli were brief (2.5 s), isolated sentences pronounced in a monotone manner whereas the stimuli presented by Lewkowicz and Hansen-Tift (2012) were 50 s monologues which for some participants were presented in an infant-directed manner (i.e., they were highly prosodic).

Because the Icelandic/English stimuli were identical in Experiments 1 and 3, we could compare the difference scores between the monolingual and bilingual participants in the two respective experiments. A 2-way Mixed ANOVA with language as the within-subject variable and task vs. no task as the between-subjects variable found a significant effect of language (English vs. Icelandic), $F(1) = 21.003$, $p < .001$, a non significant effect of task, $F(1) = 1.086$, $p < .1$, and a significant interaction, $F(1) = 7.802$, $p = .007$. This interaction supports the hypothesis that an encoding task was a critical factor in producing the differences in fixations for the English and Icelandic stimuli in Experiment 1.

5. General discussion

We found that adults devote greater attention to the source of audiovisual speech, namely a talker's mouth, when their task is to encode speech in an unfamiliar language than in an unfamiliar one. These findings complement previous results indicating that adults seek out audiovisual redundancy cues when the auditory

[3] Sample size was equated to a single language group in Experiment 1. Power analyses of the English vs. Spanish comparison in Experiment 1, which yielded a critical sample size of 22 necessary to detect an effect with 80% confidence.

[4] Sample size was equated to a single language group in Experiment 1. Power analyses of the English–Icelandic blocks in Experiment 1 yielded a critical sample size of 13 necessary to detect an effect with 80% confidence.

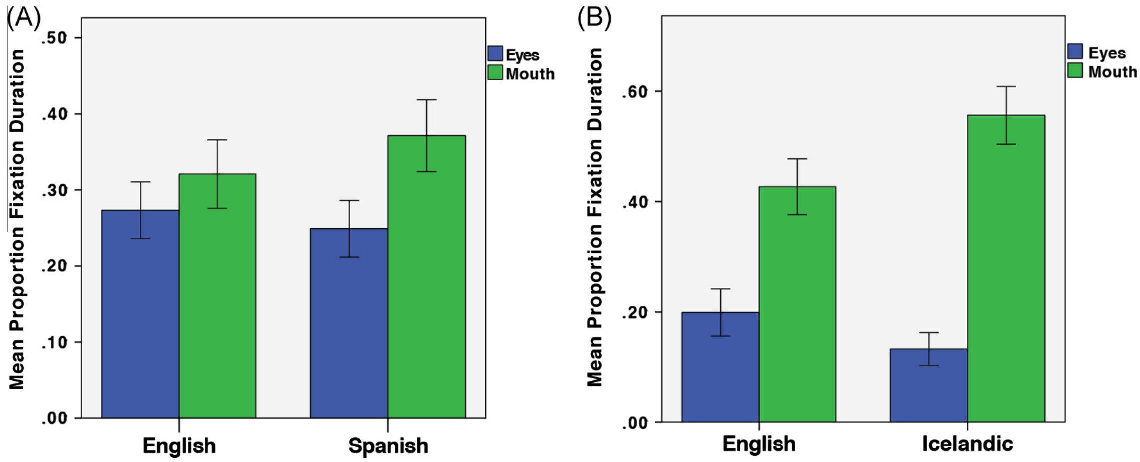

Fig. 3. Proportion of fixation duration for the eyes and mouth AOI's in Experiment 2 (A) and Experiment 3 (B). Error bars represent ±1 standard error of the mean.

signal is poor (Driver & Baylis, 1993; Lansing & McConkie, 2003; Vatikiotis-Bateson et al., 1998). Here, we show that adults also rely on audiovisual redundancy cues when dealing with high-quality speech but in an unfamiliar language. This suggests that adults possess a highly flexible system of attentional allocation that they can modulate based on the discriminability of an audiovisual speech signal, their particular speech-processing demands, and their perceptual/cognitive expertise in the particular language being uttered.

These results suggest that previous reports of enhanced mouth fixations in both adults and infants may reflect the same strategy: increased reliance on multisensory redundancy in the face of uncertainty about an audiovisual speech signal. According to this view, steadily increasing linguistic experience accounts for the shifting developmental pattern of selective attention obtained by Lewkowicz and Hansen-Tift (2012). Specifically, as infants start acquiring speech and language, initially they seek out a talker's mouth to overcome the high degree of uncertainty. Once they begin to master the various properties of their language, however, infants reduce their reliance on audiovisual redundancy cues when having to process their native/familiar language but continue to rely on these cues for the processing of a non-native/unfamiliar language. Consistent with this, evidence shows that bilingual infants, who face the serious cognitive challenge of learning two different languages, rely even more on audiovisual speech redundancy cues than do monolingual infants (Pons, Bosch, & Lewkowicz, 2015). The current findings indicate that when adults encounter unfamiliar speech, they resort to the same attentional strategy used by infants to disambiguate such speech.

Our results add to a growing body of literature demonstrating that the mouth can serve as an important source of information during audiovisual speech encoding. Nonetheless, it should be noted that attention to the mouth is not essential for integrating visual and auditory information. Audiovisual speech integration, as in the McGurk effect, can be found even at high levels of eccentricity (Paré, Richler, ten Hove, & Munhall, 2003). Thus, attention to a talker's mouth may reflect a strategy of perceptual enhancement beyond that which is absolutely necessary for integration, particularly under suboptimal encoding conditions. This enhancement may be primarily perceptual in nature, based on the higher resolution of the mouth region that comes with fixation. Alternatively, it may be primarily attentional in nature, based on additional *processing* of the fixated region. Finally, it may be due to a combination of both perceptual and attentional enhancement of the audiovisual stimulus. Additional research will be required to disentangle these potential contributing factors.

Acknowledgments

This work was supported, in part, by an NSF Award to E. Barenholtz (Grant No. #BCS-0958615) and an NIH award to D. Lewkowicz (Grant No. R01HD057116).

References

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: The roles of action, competition and social content. *The Quarterly Journal of Experimental Psychology, 61*(7), 986–998.
Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience, 2*(1), 1–13.
Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America, 116*(6), 3668–3678. http://dx.doi.org/10.1121/1.1810292.
Driver, J., & Baylis, G. C. (1993). Cross-modal negative priming and interference in selective attention. *Bulletin of the Psychonomic Society, 31*(1), 45–48.
Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews, 24*(6), 581–604.
Gat, I. B., & Keith, R. W. (1978). An effect of linguistic experience: Auditory word discrimination by native and non-native speakers of English. *International Journal of Audiology, 17*(4), 339–345.
Lansing, C. R., & McConkie, G. W. (1999). Attention to facial regions in segmental and prosodic visual speech perception tasks. *Journal of Speech, Language, and Hearing Research, 42*(3), 526–539.
Lansing, C., & McConkie, G. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception & Psychophysics, 65*(4), 536–552. http://dx.doi.org/10.3758/BF03194581.
Lecumberri, M. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America, 119*(4), 2445–2454.
Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences* (5), 1431–1436.
Love, S. A., Pollick, F. E., & Petrini, K. (2012). *Effects of experience, training and expertise on multisensory perception: Investigating the link between brain and behavior*. *Cognitive Behavioural Systems*. Springer (pp. 304–320). Springer.
Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research, 40*(3), 686–693.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748.
Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology, 56*(3), 640–662.
Middelweerd, M., & Plomp, R. (1987). The effect of speechreading on the speech-reception threshold of sentences in noise. *The Journal of the Acoustical Society of America, 82*(6), 2145–2147.
Navarra, J., Alsius, A., Velasco, I., Soto-Faraco, S., & Spence, C. (2010). Perception of audiovisual speech synchrony for native and non-native language. *Brain Research, 1323*, 84–93.
Paré, M., Richler, R. C., ten Hove, M., & Munhall, K. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception & Psychophysics, 65*(4), 553–567.
Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science, 283*(5406), 1272–1273.

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science, 26*(4), 490–498.
Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science, 17*(6), 405–409.
Rosenblum, L. D., Johnson, J. A., & Saldana, H. M. (1996). Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech, Language, and Hearing Research, 39*(6), 1159–1170.
Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour, 58*(5), 921–931.
Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212–215.
Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica, 36*(4–5), 314–331.
Van Wijngaarden, S. J., Steeneken, H., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America, 111*(4), 1906–1916.
Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., & Munhall, K. (1998). Eye movement of perceivers during audiovisualspeech perception. *Perception & Psychophysics, 60*(6), 926–940.