Psychon Bull Rev (2012) 19:66–72
DOI 10.3758/s13423-011-0176-8

BRIEF REPORT

Can you McGurk yourself? Self-face and self-voice in audiovisual speech

Christopher Aruffo · David I. Shore

Published online: 28 October 2011

Abstract We are constantly exposed to our own face and voice, and we identify our own faces and voices as familiar. However, the influence of self-identity upon self-speech perception is still uncertain. Speech perception is a synthesis of both auditory and visual inputs; although we hear our own voice when we speak, we rarely see the dynamic movements of our own face. If visual speech and identity are processed independently, no processing advantage would obtain in viewing one's own highly familiar face. In the present experiment, the relative contributions of facial and vocal inputs to speech perception were evaluated with an audiovisual illusion. Our results indicate that auditory self-speech conveys a processing advantage, whereas visual self-speech does not. The data thereby support a model of visual speech as dynamic movement processed separately from speaker recognition.

Keywords Speech perception · Face perception and recognition · Psycholinguistics · Self-speech

Introduction

We are constantly exposed to our own face and voice. We frequently see our face in the mirror, in photographs, and in video recordings. Every time we open our mouth to speak, we hear our voice internally. But we rarely sit in front of a mirror and watch our own lips as we talk, and an external recording of our own voice is markedly different from the sound we know. We may be constantly exposed to our self-identity, but does this exposure translate into superior processing for self-speech?

People identify their own faces and voices as familiar (Keyes, Brady, Reilly, & Foxe, 2009; Nakamura et al., 2001). Behaviorally, self-face is recognized more quickly and accurately than familiar or nonself faces (Keyes & Brady, 2010), and contemporary listeners discriminate between *self*, *familiar*, and *other* voices with near-ceiling accuracy (Rosa, Lassonde, Pinard, Keenan, & Belin, 2008). Neurologically, visual and auditory self-identification activate regions related exclusively to self-processing (Kaplan, Aziz-Zadeh, Uddin, & Iacoboni, 2008). It is yet undetermined whether the special processing observed for self-identity influences the perception of external self-speech.

The influence of identity recognition upon speech perception is still uncertain. Identification and speech processing may be disassociated both for voice (Relander & Rämä, 2009) and for face (Campbell, Landis, & Regard, 1986), but recognition and articulation do appear to exert some influence upon each other. For example, characteristic speech production is integral to a speaker's vocal identity (Sheffert, Pisoni, Fellowes, & Remez, 2002); it therefore stands to reason that a familiar voice automatically increases speech intelligibility (Nygaard & Pisoni, 1998), but it is not known whether self-voice, although familiar, will automatically facilitate speech comprehension. Similarly, judgments of facial identity incorporate facial movement (Knappmeyer, Thornton, & Bülthoff, 2003), and familiarization with a speaking face improves lipreading comprehension (Lander & Davies, 2008), but speech and identity may be separate judgments arising from the same

C. Aruffo · D. I. Shore
McMaster University,
Hamilton, ON, Canada

C. Aruffo (✉)
Department of Psychology, Neuroscience, & Behaviour,
1280 Main Street West,
Hamilton, ON L8S 4K1, Canada
e-mail: aruffocc@mcmaster.ca

facial input (Bruce & Young, 1986). It may be that exposure impresses dynamic facial movements upon visual memory and these memories become associated with auditory constructs, such that their activation contributes to improved speech comprehension (von Kriegstein et al., 2008) by generating stronger representations during speech perception (van Wassenhove, Grant, & Poeppel, 2005). It is therefore likely that a face whose features are familiar, but whose lip movements have not been studied, will not automatically improve speech intelligibility.

The relative contribution of facial and vocal inputs to speech perception can be evaluated with the *McGurk effect*. The McGurk effect is an audiovisual illusion in which incongruent visual and auditory speech signals are integrated (McGurk & MacDonald, 1976); for example, an auditory /ba/ dubbed onto a visual /ga/ is perceived as /da/ or /la/. Observers are unable to ignore the illusory percept, which demonstrates the mandatory influence of vision on auditory speech perception and makes the McGurk illusion ideal for testing audiovisual speech.

The illusion may take one of two forms: a “blend” illusion, in which auditory /ba/ and visual /ga/ are *blended* into a percept such as /da/, or a “combination” illusion, in which auditory /ga/ and visual /ba/ are *combined* into a percept such as /bga/ (MacDonald & McGurk, 1978). Combinations and blends produce different hemifield responses (Diesch, 1995), and self-voice recognition exhibits a right-hemisphere advantage (Hughes & Nicholson, 2010; Rosa et al., 2008). To accommodate the possibility of an interaction between self-speech and type of illusion, both blend and combination illusions were included in our experimental design.

The present experiment focused on self-speech. It did not include other familiar speakers, because familiar identities may not integrate when mismatched (Walker, Bruce, & O’Malley, 1995). Multiple speakers were presented to each participant, but this was not expected to have an influence on audiovisual perception. That is, switching speaker identity from trial to trial has been observed to influence perception of either auditory or visual speech (Mullenix, Pisoni, & Martin, 1989; Yakel, Rosenblum, & Fortier, 2000), but identity variation does not appear to influence the audiovisual McGurk effect (Rosenblum & Yakel, 2001). Thus, the present experiment was designed to compare self-speech with nonself-speech.

To test the effect of self-speech, we recorded participants speaking McGurk stimuli and presented them with these stimuli. Stimuli were nonwords to avoid lexical effects (Brancazio, 2004). Disyllabic stimuli were used because an initial vowel would allow vocal identification prior to presentation of the target consonant (Beauchemin et al., 2006). Visual speech activity prior to presentation of a speech target can facilitate perception of the target (van Wassenhove et al., 2005). If self-identity facilitates speech processing, disyllabic stimuli should be sufficient to demonstrate an effect of self-speech.

Each participant was exposed to three different block types. In *matched* blocks, the face and voice on each trial came from the same individual, with half of these trials containing a self-recording. A comparison of the quantity of perceived illusions for trials containing self-recordings versus trials containing nonself-recordings provides a measure of the impact of self-identity on speech integration. If a familiar identity reduces the illusion only when mismatched (Walker et al., 1995), the illusion will be equally strong for self and nonself trials. There were two types of *mismatch* blocks: In mismatched-*other* blocks, all of the faces and voices were from nonself individuals; in mismatched-*self* blocks, each trial contained either self-face or self-voice dubbed into nonself-voice or nonself-face. Here, the critical comparison concerns the quantity of perceived illusions when self is not present (mismatched-other) versus the quantity of perceived illusions when either self-face or self-voice is present (mismatched-self). On the basis of the observation that we rarely see ourselves talking but hear ourselves constantly, we expect the auditory channel to be more reliable. Therefore, because a more reliable auditory channel reduces the strength of the McGurk illusion (Massaro & Cohen, 1990), we expect a reduced quantity of illusions for trials featuring self-voice.

For each model participant, we recruited two individuals to observe the same self-trials as that model. Each of these control participants thus acted as a surrogate “self.” We attribute any performance difference between these two groups of participants as supporting the unique role of self-identity in speech perception.

Method

Participants

There were 19 *model* participants and 38 *control* participants. Model participants (7 male, 12 female; age, 18–45 years) provided stimulus recordings. Control participants’ (8 male, 30 female; age, 17–26 years) data were collected from 15 of the 19 models (5 male, 10 female). All participants gave informed consent to the procedures. All procedures complied with the tricouncil ethics procedures in Canada as approved by the McMaster Research Ethics Board. All participants were McMaster University students who received course credit for their participation.

Stimuli

Each model recorded five repetitions of six disyllables: /aba/, /ada/, /aga/, /ala/, /aða/, and /abga/. Models were instructed to articulate each "as though there were a comma between each syllable." When recorded, models were seated before a plain beige background in a sound-attenuated room. A digital video camera (JVC GZ-MG37U) and wireless lapel microphone (Shure PG185) were used. Two 60-W lamps were placed at 45° angles to the participant's body, situated approximately 2 ft away from the participant and on a vertical level with the participant's face, thus rendering the face clearly and fully visible. The video camera was also on a vertical level with the face. Video was recorded in standard 4:3 aspect ratio, framed to include the entire face, to the camera's internal hard drive in MPEG-2 format (720 × 480 pixels, 8.5 Mbps). Audio was digitally recorded from the microphone in lossless WAV format (16-bit depth, 44,100 samples/s). The volume of each was normalized using Adobe Audition 3. The audio and video of each session were combined and synchronized to within 6-ms accuracy. The resulting audiovisual streams were cut into 4-s-long segments. Each segment displayed a participant's face gazing silently for approximately 3 s and then speaking one complete disyllable. All audiovisual editing was performed with Adobe Premiere CS4.

Three classes of disyllables were created from these audiovisual segments: *congruent*, *blend*, and *combination* stimuli. Each congruent stimulus presented the same auditory and visual disyllable (/aba/, /ada/, /aga/, /ala/, /aða/, abga/). Each blend stimulus dubbed auditory /aba/ onto visual /aga/; the illusory percept of this stimulus is a phoneme not physically present, such as /ada/, /ala/, or /aða/. Each combination stimulus dubbed auditory /aga/ onto visual / aba/; its illusory percept was a combination of both phonemes, or /abga/. The six congruent disyllables thereby mirrored the most common percepts induced by blend and combination stimuli.

Matched stimuli presented a face and voice belonging to the same speaker. Because each participant had recorded five repetitions of each disyllable, this resulted in 30 congruent stimuli (6 disyllables × 5 repetitions). To create illusory disyllables, the five repetitions of /aba/ and /aga/ were dubbed onto each other, producing five blends and five combinations. A total of 40 matched stimuli were thus produced for each model.

Mismatched stimuli presented faces and voices belonging to different speakers. For each disyllable, each model's voice was dubbed onto every other model's face. Each mismatched face, therefore, generated 108 congruent stimuli (6 disyllables × 18 voices), 18 blends, and 18 combinations. In each stimulus, the consonant release was synchronized within an accuracy of 12 ms, and the peak intensity of the initial vowel within an accuracy of 16 ms. Differences in speaking rate between models were accommodated by increasing or decreasing the period of silence between syllables in the auditory track to match the (unaltered) video track. A total of 144 mismatched stimuli were thus produced for each model.

A grand total of 3,496 unique audiovisual stimuli were created. In each stimulus, disyllables were either *congruent* or *illusory*, and identity was either *matched* or *mismatched*. Five hundred seventy stimuli were *congruent matched* (6 disyllables × 5 repetitions × 19 models); 190 were *illusory matched* (2 × 5 × 19); 2,052 were *congruent mismatched* (6 disyllables × 19 faces × 18 voices); and 684 were *illusory mismatched* (2 × 19 × 18). Not all stimuli were presented to every participant. These stimuli formed a pool from which the experimental procedure drew.

Procedure

The testing session was conducted in a sound-attenuated room using a 15-in. widescreen laptop computer (Gateway W6501) at 1,280 × 800 resolution. Display brightness was set to maximum. All display elements other than the experimental interface were blanked from view. Audio was played from the computer's speakers situated directly below the screen. Volume was set to maximum; all participants were played a single stimulus of their own face and voice speaking /ala/ to verify that the volume was neither inaudible nor uncomfortably loud.

No participant saw a familiar nonself speaker. Prior to testing, participants were shown silent segments of each speaker articulating /ala/ and indicated any familiarity with that speaker. When a speaker was identified as familiar, this identification operated as a constraint during the testing procedure, such that stimuli featuring that speaker would not be selected for presentation.

Controls were each assigned to a particular model as their "self." That is, each of the 38 controls was assigned to a specific "self," allowing each of the 19 models to be presented as "self" in two different control sessions. Four of the models did not provide data; however, their absence was not anticipated at the commencement of the testing process, so their stimuli were included in the procedure. Matching sexual identity was not expected to be a factor (Green, Kuhl, Meltzoff, & Stevens, 1991) and was not controlled. By assigning a "self" to each control, a control would be presented with blocks of stimuli concordant with those presented to a particular model in that model's experimental session. Thus, controls were presented with stimuli representing "self-match," "self-face," and "self-voice"; however, it must be understood that these stimuli featured the particular model assigned to that control. Controls did not see their own face or voice.

Stimuli were organized by identity into three block types: *matched*, *mismatched-self*, and *mismatched-nonself*. Each block presented 48 trials. *Matched* blocks featured 24 self-speech and 24 nonself-speech; *mismatched-self* featured 24 self-face and 24 self-voice; *mismatched-nonself* featured no self-stimuli. Within each block, stimuli were selected randomly at runtime from the available pool, with two constraints: No two identical stimuli were presented consecutively, and no nonself stimulus featured a familiar identity. Each unfamiliar nonself speakers was presented with equivalent frequency across all blocks. Each type of block was presented twice, making a total of six blocks and a grand total of 288 trials.

Each set of 24 trials comprised nine blend and nine combination stimuli plus one of each congruent disyllable. Experiments featuring the McGurk effect typically feature only blends and combinations; congruent stimuli, if used, are tested beforehand to confirm a speaker's intelligibility. However, the present experiment was different in that models could, potentially, remember what they had recorded. If models had recorded only /aba/ and /aga/, they might, when tested, reject percepts they knew they had not spoken. Alternatively, having recorded six disyllables, models tested only with blends and combinations might wonder why they were not hearing all six. Including congruent stimuli in the experimental design provided controls for each model's intelligibility while reassuring models that all six disyllables were present and possible.

Participants watched single tokens and verbally repeated what they heard. Participants were instructed to maintain eye gaze on the screen and were informed that their looking away would produce invalid results. Verbal free responses were solicited because manual closed-response is known to strengthen the McGurk illusion (Colin, Radeau, & Deltenre, 2005). In a pilot experiment, responses had been recorded by an experimenter with paper and pen; however, these responses were not always heard correctly. In the present experiment, participants' responses were recorded using an omnidirectional microphone (Shure SM58); additionally, to guard against mishearings, participants reported their own verbal response on each trial by clicking on one of five buttons reading "ABA," "ADA," "AGA," "ALA," "ATHA," and "OBGA," or by typing a free response. To avoid any influence of visible letters upon an illusory percept, the buttons were hidden while each stimulus was presented and reappeared 3 s after the stimulus had ended. Participants were instructed that this report was meant not to be a "second chance" but to be an accurate record of what they had said. The input given to the computer, as an accurate representation of verbal responses, generated the data analyzed here.

Results

Results were measured as the proportion of trials causing *integrated* (illusory) percepts. A percept was *nonintegrated* when either the auditory or the visual channel was reported alone (/aba/ or /aga/) and *integrated* when neither the auditory nor the visual channel was accurately reported (e.g., /ala/, /ada/). Congruent trials served as controls for intelligibility. Of the congruent trials, 92% were reported correctly across all identity conditions and did not require further analysis.

No effect of cross-sex conflict between face and voice was expected. Nonetheless, to evaluate this factor, a 2 × 2 × 3 ANOVA (group, sex conflict, identity) was performed on mismatched-identity trials, using the between-subjects factor of *group* (model, control) and the within-subjects factors of *sex conflict* (conflict, no conflict) and *identity* (self-face, self-voice, and nonself-mismatch). No effect of sex conflict was found, $F(1, 51) = 0.11$, n.s., and no interactions were observed. As such, the factor of sex was not considered further.

The presence of self-speech reduced overall susceptibility to the McGurk illusion. A 2 × 2 × 5 ANOVA was performed (group, disyllable, identity) on all data; *disyllable* represented either blends or combinations, and *identity* included five levels (nonself-match, self-match, nonself-mismatch, self-face, and self-voice). The model group produced fewer integrated responses than did controls, $F(1, 51) = 15.26$, $p < .001$, and the effect of identity condition differed for the two groups, $F(4, 204) = 5.28$, $p < .001$ (for the interaction of group and identity condition, compare Fig. 1a and b). No effect of disyllable was observed, and the three-way interaction was not significant. The factor of disyllable was therefore removed in analyzing matched and mismatched blocks.

Self-speech influenced both matched and mismatched speech. Two separate ANOVAs (group, identity) were performed to confirm the interaction of self-speech and identity in matched trials (2 × 2), $F(1, 51) = 6.43$, $p = .014$, and in mismatched trials (2 × 3), $F(2, 102) = 6.456$, $p = .002$. Four planned *t*-tests were performed ($\alpha = .05$). In matched blocks, self-speech integrated to a lesser degree than nonself-speech, $t(14) = 4.006$, $p = .001$. When mismatched, self-voice supported fewer integrations than did nonself, $t(14) = 2.764$, $p = .015$, whereas self-face did not differ from nonself-face, $t(14) = .647$, n.s.. Finally, self-voice supported fewer illusions than did self-face, $t(14) = 3.148$, $p = .007$, in mismatched-self blocks. Among all illusory trials, 40.0% were reported as auditory (/aba/) and 1.5% as visual (/aga/). Thus, self-speech reduces the magnitude of the McGurk effect, which reduction appears to be caused primarily by self-voice.

We also analyzed blends and combinations separately on the basis of our a priori assumption that these trial types may produce different results (cf. Diesch, 1995). Simple-effects analyses confirmed an effect of identity on the proportion of integrated responses on blend trials,

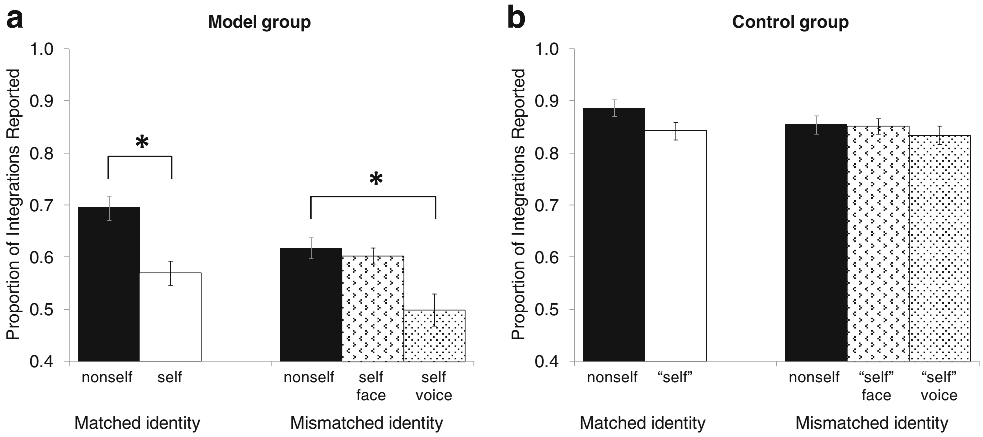

Fig. 1 Proportion of illusions perceived under different identity conditions, separated by participant group. Error bars represent standard error of mean corrected for within-subject comparisons. **a** Model group. **b**: Controls

F(4, 56) = 6.484, *p* < .001. Combination trials showed no such effect, *F*(4, 56) = .856, n.s.; although the same pattern seen with the overall data and with the blends was weakly suggested, the differences were not nearly as striking (compare Fig. 2a and b). Two-tailed planned paired-samples *t*-tests, performed on blends, show fewer integrations supported by self-speech than by nonself-speech, *t*(14) = 3.35, *p* = .005; mismatched blends show marginal differences between nonself- and self-voice, *t*(14) = 2.102, *p* = .054, or self-face and self-voice, *t*(14) = 2.142, *p* = .050.

Discussion

Processing expertise for audiovisual self-speech was tested using the McGurk effect. Self-face, self-voice, and audiovisual self-speech were compared with more traditional conditions that did not include self-voice or self-face. Fewer integrated stimuli were perceived as illusory when self-speech was viewed, as compared with nonself speech. Examining the relative roles of self-voice and self-face versus mismatched-nonself showed fewer integrations for self-voice only. None of these effects were observed when the same stimuli were presented to control participants. We interpret this effect as a greater processing expertise for self-voice than for self-face in the integration of self-speech stimuli.

Auditory self-speech conveys a processing advantage. The advantage is observed here by heavier weighting being given to the auditory channel, and not to the visual channel, when audiovisual speech was integrated. In mismatched conditions, self-voice weakened the illusion, whereas self-face did not. In matched conditions, self-speech supported fewer illusions, and responses were mainly auditory (/aba/); the auditory response cannot be attributed to attentional selection or signal clarity, since both channels were from the same source and free of noise or degradation. The result may be interpreted as greater weight to the auditory channel. That is, auditory self-speech was perceived as more reliable than visual self-speech and, thus, was weighted more heavily. The reduction in illusion strength across all conditions for the model group may be accounted for by this emphasis on self-voice. That is, the greater reliability of the voice in the self-voice condition may carry over to a more general emphasis on the voice and a subsequent overall reduction in illusion strength. This conclusion must remain speculative at this time, since we did not predict or expect an overall reduction in the strength of the McGurk illusion in the presence of self-speech.

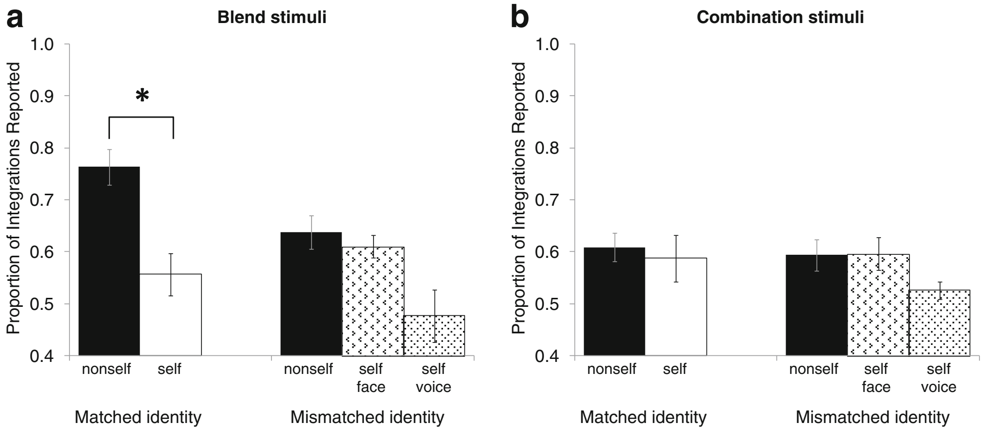

Fig. 2 Proportion of illusions perceived by model group under different identity conditions, separated by disyllable type. Error bars represent standard error of the mean corrected for within-subjects comparisons. **a** Blend trials. **b** Combination trials

Self-face does not provide an advantage to visual speech processing. Self-face trials supported the same number of integrations as nonself-mismatched trials. Because self-face is familiar, this result supports the dual-channel model of face perception (Bruce & Young, 1986) by indicating that improved lipreading skill is not automatically conferred by familiarity with a speaker's facial features. Rather, a listener must watch a person speaking to become familiar with that speaker's visual speech.

Self-identification may exert some influence on audiovisual speech integration. This influence is implied by the difference between blend and combination trials. Although combination trials showed an overall reduction in integrated percepts, as compared with controls, this reduction may be explained by the already-mentioned adaptation to the auditory channel. Because combination illusions are better perceived in the right visual field (Diesch, 1995) and self-speech is better recognized in the right hemisphere of the brain (Hughes & Nicholson, 2010), it may be supposed that self-recognition did not affect combination trials, or affected them too weakly to be observed within the present experimental design. The difference observed between identity conditions for blend trials might therefore be partially attributed to self-identification.

Self-identification does not prevent integration of audiovisual speech. Self-face integrated with nonself-voice in the same proportion as any two unknown speakers. Although self-voice supported fewer integrations than did self-face, integration was nonetheless greater than floor performance, indicating that self-voice did integrate with nonself faces to some degree. Therefore, the difference between self-face and self-voice might most accurately be explained by optimal weighting of the sensory input (cf. Ernst & Banks, 2002).

Self-face will integrate with an unknown voice. This result argues against an automatic influence of facial identity upon facial speech, as suggested by Walker et al. (1995). In that experiment, familiar speakers reduced the illusion only when face and voice were mismatched. If that result had been due to facial speech, optimal weighting toward the visual channel would have occurred, and participants would have reported visual percepts. However, their participants reported auditory percepts. It may be that when both face and voice are separately identifiable, a mismatched stimulus may be disunified into separate channels, allowing attentional selection to the auditory response. Conversely, the integration of matched stimuli observed in that experiment can be parsimoniously explained: If facial and vocal speech are equally reliable, neither would exert such weight as to produce a nonintegrated percept.

Although there is still much to be discovered about the effects of familiarity and self-speech on audiovisual speech perception, the present data argue that familiar facial features do not automatically contribute to audiovisual speech processing.

Conclusion

Auditory self-speech resisted integration with visual speech from other speakers, supporting the claim of a processing advantage for auditory self-speech. Self-face did not resist integration with unknown voices, supporting the supposition that speakers are familiar with features of their own voice but less so with the dynamic features of their own face speaking. Furthermore, the lack of an advantage for self-face stimuli refutes the claim that familiar facial features always provide an advantage, confirming instead that familiarity with the specific dynamic stimulus may be required. This result has implications for how we process information about ourselves and about others with whom we are familiar.

Author note and acknowledgements These data were presented in partial fulfillment of an M.Sc. degree to the first author. We would like to thank David Feinberg and Karin Humphreys for supervision and support throughout this project, and Kyla Baird, who assisted in collecting data. We also acknowledge the Natural Sciences and Engineering Research Council of Canada who supported D.I.S. through a Discovery Grant, and the Canadian Foundation for Innovation and Ontario Innovation Trust for infrastructure funding. These data were previously presented at IMRF 2011 in Japan, imrf.info/2011/480.

References

Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P., et al. (2006). Electrophysiological markers of voice familiarity. *European Journal of Neuroscience, 23,* 3081–3086. doi:10.1111/j.1460-9568.2006.04856.x

Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 445–463. doi:10.1037/0096-1523.30.3.445

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77,* 305–327.

Campbell, R., Landis, T., & Regard, M. (1986). Face recognition and lipreading. *Brain, 109,* 509–521. doi:10.1093/brain/109.3.509

Colin, C., Radeau, M., & Deltenre, P. (2005). Top-down and bottom-up modulation of audiovisual integration in speech. *European Journal of Cognitive Psychology, 17,* 541–560. doi:10.1080/09541440440000168

Diesch, E. (1995). Left and right hemifield advantages of fusions and combinations in audiovisual speech perception. *Quarterly Journal of Experimental Psychology, 48A,* 320–333.

Ernst, M., & Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433. doi:10.1038/415429a

Green, K., Kuhl, P., Meltzoff, A., & Stevens, E. (1991). Integrating speech information across talkers, sex, and sensory modality:

Female faces and male voices in the McGurk effect. *Perception & Psychophysics, 50,* 524–536.

Hughes, S., & Nicholson, S. (2010). The processing of auditory and visual recognition of self-stimuli. *Consciousness and Cognition, 19,* 1124–1134. doi:10.1016/j.concog.2010.03.001

Kaplan, J., Aziz-Zadeh, L., Uddin, L., & Iacoboni, M. (2008). The self across the senses: An fMRI study of self-face and self-voice recognition. *Social Cognitive and Affective Neuroscience, 3,* 218–223. doi:10.1093/scan/nsn014

Keyes, H., & Brady, N. (2010). Self-face recognition is characterized by "bilateral gain" and by faster, more accurate performance which persists when faces are inverted. *Quarterly Journal of Experimental Psychology, 63,* 840–847.

Keyes, H., Brady, N., Reilly, R., & Foxe, J. (2009). My face or yours? Event-related potential correlates of self-face processing. *Brain and Cognition, 72,* 244–254. doi:10.1016/j.bandc.2009.09.006

Knappmeyer, B., Thornton, I., & Bülthoff, H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research, 43,* 1921–1936. doi:10.1016/S0042-6989(03)00236-0

Lander, K., & Davies, R. (2008). Does face familiarity influence speechreadability? *Quarterly Journal of Experimental Psychology, 61,* 961–967. doi:10.1080/17470210801908476

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

Massaro, D., & Cohen, M. (1990). Perception of synthesized audible and visible speech. *Psychological Science, 1,* 55–63. doi:10.1111/j.1467-9280.1990.tb00068.x

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748. doi:10.1038/264746a0

Mullennix, J., Pisoni, D., & Martin, C. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85,* 365–378. doi:10.1121/1.397688

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., et al. (2001). Neural substrates for recognition of familiar voices: A PET study. *Neuropsychologia, 39,* 1047–1054.

Nygaard, L., & Pisoni, D. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60,* 355–376.

Relander, K., & Rämä, P. (2009). Separate neural processes for retrieval of voice identity and word content in working memory. *Brain Research, 1252,* 143–151. doi:10.1016/j.brainres.2008.11.050

Rosa, C., Lassonde, M., Pinard, C., Keenan, J., & Belin, P. (2008). Investigations of hemispheric specialization of self-voice recognition. *Brain and Cognition, 68,* 204–214. doi:10.1016/j.bandc.2008.04.007

Rosenblum, L., & Yakel, D. (2001). The McGurk effect from single and mixed speaker stimuli. *Acoustics Research Letters Online, 2,* 67–72. doi:10.1121/1.1366356

Sheffert, S., Pisoni, D., Fellowes, J., & Remez, R. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1447–1469. doi:10.1037//0096-1523.28.6.1447

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

von Kriegstein, K., Dogan, Ö., Grüter, M., Giraud, A., Kell, C., Grüter, T., et al. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences, 105,* 6747–6752.

Walker, S., Bruce, V., & O'Malley, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. *Perception & Psychophysics, 57,* 1124–1133.

Yakel, D., Rosenblum, L., & Fortier, M. (2000). Effects of talker variability on speechreading. *Perception & Psychophysics, 62,* 1405–1412.