Attention, Perception, & Psychophysics
2010, 72 (6), 1450-1454
doi:10.3758/APP.72.6.1450

A comparison of the McGurk effect for spoken and sung syllables

Lena Quinto and William Forde Thompson
Macquarie University, Sydney, New South Wales, Australia

Frank A. Russo
Ryerson University, Toronto, Ontario, Canada

and

Sandra E. Trehub
University of Toronto, Mississauga, Ontario, Canada

The importance of visual cues in speech perception is illustrated by the McGurk effect, whereby a speaker's facial movements affect speech perception. The goal of the present study was to evaluate whether the McGurk effect is also observed for sung syllables. Participants heard and saw sung instances of the syllables /ba/ and /ga/ and then judged the syllable they perceived. Audio–visual stimuli were congruent or incongruent (e.g., auditory /ba/ presented with visual /ga/). The stimuli were presented as spoken, sung in an ascending and descending triad (C E G G E C), and sung in an ascending and descending triad that returned to a semitone above the tonic (C E G G E C♯). Results revealed no differences in the proportion of fusion responses between spoken and sung conditions confirming that cross-modal phonemic information is integrated similarly in speech and song.

Perhaps the best-known example of visual influences on speech perception is the McGurk effect (McGurk & MacDonald, 1976). When an audible syllable is presented simultaneously with the facial movements used to produce a different syllable, the perception of a third syllable can result. For example, when the sound /ba/ is presented with the visible articulation of /ga/, often /da/ is perceived. This illusion demonstrates the fusion of auditory and visual cues.

The McGurk effect does not depend on precise synchrony, spatial resolution, or specific gaze patterns. Auditory stimuli can be delayed by as much as 250 msec or advanced by 60 msec relative to visual stimuli before the McGurk effect is attenuated (Massaro, Cohen, & Smeele, 1996; Munhall, Gribble, Sacco, & Ward, 1996). The effect is also observed under conditions of decreasing spatial resolution, although it becomes less frequent as the dominance of auditory cues increases (MacDonald, Andersen, & Bachmann, 2000). Participants' gaze needs to be displaced by at least 10º–20º from the mouth area before the McGurk effect is weakened (Paré, Richler, ten Hove, & Munhall, 2003). Even when participants observe their own silent syllable articulation in a mirror while hearing an incongruent auditory syllable, they report neither the syllable they articulated nor the syllable they heard, but rather a fused syllable that combines elements of both (Sams, Möttönen, & Sihvonen, 2005). A number of researchers have proposed theories of how auditory and visual input are integrated to create a unified percept. According to *optimal integration theory*, the relative influence of each modality depends on its perceived reliability in conveying information (Alais & Burr, 2004; Massaro, 2004).

Visual information also affects the perception of music. For example, body movements can change the perception of musical tension and expressiveness (Dahl & Friberg, 2007; Davidson, 1993; Vines, Krumhansl, Wanderley, & Levitin, 2006), tone duration (Schutz & Lipscomb, 2007), and plucked versus bowed judgments (Saldaña & Rosenblum, 1993). Facial cues can influence the perception of happiness and sadness conveyed by music (Thompson, Graham, & Russo, 2005; Thompson, Russo, & Quinto, 2008) and the size of melodic intervals (Thompson, Russo, & Livingstone, 2010). The above studies reflect a simple compromise between auditory and visual cues rather than a wholly independent fused percept. The question addressed in the present study is whether a fused percept (i.e., the McGurk effect) occurs for sung syllables as it does for spoken syllables.

Although spoken and sung syllables are similar in some respects, there are several reasons for expecting that perceptual effects might diverge for these two contexts. First, behavioral, neuropsychological, and ERP evidence suggests that musical and linguistic abilities are dissociable (Besson, Faïta, Peretz, Bonnel, & Requin, 1998; Bonnel, Faïta, Peretz, & Besson, 2001; Peretz & Coltheart, 2003), whereas other evidence from brain imaging points to se-

W. F. Thompson, bill.thompson@mq.edu.au

lective overlap in the processing of music and speech (for reviews, see Patel, 2003, 2009). As Patel (2009) pointed out, processing demands for language and music are comparable for some attributes (i.e., syntactic structure, melodic and prosodic pitch contour) but not for others (e.g., polyphony in music, semantic content in language).

Second, pitch operates differently in the two domains and may lead to differences in where attention is directed. Singing typically involves a succession of discrete pitches, in contrast to the pitch glides that are characteristic of speech ('t Hart, Collier, & Cohen, 1990). In song, prescribed pitches within a learned tonal framework are used, creating expectations about the pitches that follow (Krumhansl, 1990). Because pitch movement in speech is much less constrained than it is in music, it is difficult to encode in detail, encouraging attentional resources to be directed toward linguistic content. In speech, linguistic content is the focus of attention, whereas in music, pitch and temporal relations are critical. In a sung context, linguistic content may be secondary to pitch and temporal structure. Some studies have revealed that vowel intelligibility (Gregg & Scherer, 2006; Hollien, Mendes-Schwartz, & Nielsen, 2000) and syllable articulation (Collister & Huron, 2008) are less clear in sung than in spoken contexts.

Third, the quality and intelligibility of the visual information associated with singing and speech are different. In speech, when auditory information is not clear, visual cues can provide additional information such as place of articulation (Neely, 1956) and lead to increased intelligibility of speech in noise (Sumby & Pollack, 1954). However, these findings need not extend to singing. Hidalgo-Barnes and Massaro (2007) observed only modest benefits to the identification of song lyrics when visual information was available. When asked to identify lyrics, participants were correct 28% of the time in the auditory condition and 33% in the audio–visual condition. This increase in intelligibility is modest in comparison with that typically observed for spoken stimuli (but see Jesse & Massaro, 2010).

The visual cues associated with singing may not be as informative as those observed in speech. There are notable differences between speech and singing with respect to the magnitude and quality of facial actions. In speech, jaw and lip movements are relatively small (Ostry & Munhall, 1994; Smith, 1992). In singing, larger jaw openings are used for producing higher pitches, preserving amplitude, and articulating sung vowels (Sundberg, 1977). The movements required to reach and maintain pitches in singing contribute to the lower intelligibility of phonemic information in this domain and may influence how integration of audio–visual information occurs.

In short, there are three primary motivations for comparing the McGurk effect under sung and spoken conditions. First, such studies contribute to research on the nature and extent of overlap in mechanisms underlying music and speech (see, e.g., Patel, 2008, for a review). Second, they contribute to the understanding of phoneme perception in sung contexts (e.g., Collister & Huron, 2008; Sundberg, 1977). Third, they contribute to research on processes of audio–visual integration and the contexts that influence those processes.

In the present investigation, participants were presented with spoken or sung syllables as they watched displays of visually articulated syllables that were congruent or incongruent with the auditory materials. An out-of-key condition was included to determine whether integration is affected by expectancy violation. Previous research indicates that unexpected notes or chords can in some instances lead to increased processing of concurrently presented nonmusical material (Bigand, Tillmann, Poulin, D'Adamo, & Madurell, 2001). Expectancy violation could therefore affect integration by drawing attention to the verbal material that accompanies the unexpected pitch. More generally, if audio–visual integration operates similarly in spoken and sung contexts, then incongruent audio–visual conditions should result in the perception of fused syllables in both spoken and sung contexts, with similar rates of occurrence. If audio–visual integration operates differently in sung and spoken contexts, we should observe significant overall differences in the McGurk effect for these contexts.

METHOD

Participants

The participants were 28 students in an introductory psychology course (18–22 years of age; $M = 19.07$, $SD = 1.18$); all received partial course credit or token payment for their participation. All reported normal hearing and vision. Participants had an average of 2.89 years of musical training.

Stimuli

A highly trained vocalist was video-recorded speaking and singing in a professional recording studio using a digital Sony Handycam DCRHC52 and an Avantones CR-14 microphone. Audio and visual channels were recorded separately for subsequent resynchronization. Audio stimuli were recorded with ProTools (7.2, Digidesign, Daly City, CA). The audio and visual components were synchronized by aligning the timing of the consonant burst in the audio and visual signals with Final Cut Pro (5.1, Apple, Cupertino, CA). Nine stimuli were created: three audio–visual conditions (congruent /ba/, congruent /ga/, and audio /ba/ combined with visual /ga/) for each of three expressive conditions (spoken, sung in key, and sung out of key). For the out-of-key condition, the video recording of the in-key condition was used to prevent the influence of visual cues associated with deliberate off-key singing.

The stimuli for each trial consisted of six spoken or sung syllables. During recording, the vocalist expressed the syllables /ba/ and /ga/ six times to produce either a spoken or sung trial. She was also required to synchronize the onset of the syllables with a click track set to 70 beats per min presented through headphones. The use of the click track ensured that recordings were consistent in timing with one another, allowing us to synchronize the audio track for one syllable with the video track for a different syllable. For the spoken condition, the vocalist was asked to utter the six syllables in a normal voice. For the sung condition, she was first presented with a six-note melody through headphones and then sang that melody using one of the two syllables. For the in-key condition, the vocalist sang an ascending–descending triad (C E G G E C). For the out-of-key condition, the vocalist sang an ascending–descending triad that returned to the semitone above the tonic (C E G G E C♯). The vocalist sang with breaks between notes at a rate of one syllable per beat (~857 msec per syllable).

Mean fundamental frequencies ($f0$) for spoken and sung syllables, as determined by Praat software (Boersma & Weenink, 2007), are displayed in Figure 1. The range of $f0$ values was approximately 3 Hz for spoken stimuli and approximately 140 Hz for sung stimuli.

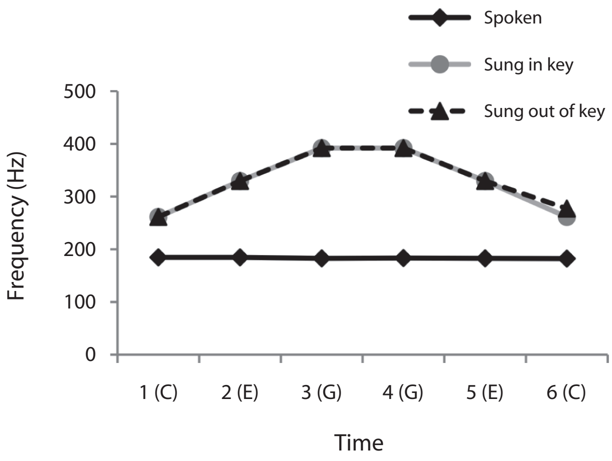

Figure 1. Changes in fundamental frequency across time for the three conditions. The final note was C♯ rather than C in the out-of-key condition.

The click track ensured that onset-to-onset times were identical for spoken and sung stimuli. However, because stimuli were produced naturally, other attributes of spoken and sung syllables were not controlled.

Procedure

Testing was conducted in a sound-attenuating booth. All participants judged the audio–visual stimuli. Half of the participants (n = 14) then judged auditory and visual unimodal stimuli in order to provide a baseline assessment. Unimodal judgments were obtained last to avoid cuing participants to the experimental conditions. Participants were told that the study was designed to explore the perception of lyrics in music. Their task was to name the last syllable that they heard in a series of six syllables. Visual stimuli were presented on a 17-in. Viewsonic VX724 LCD monitor. Auditory stimuli were presented binaurally through Sennheiser HD280 headphones at a comfortable listening level—approximately 60 dB. Participants were asked to watch the monitor and attend carefully to the spoken or sung syllables so that they could identify the final syllable that they heard. After the presentation of every stimulus, five options appeared on the monitor: *ba*, *da*, *ga*, *tha*, and *other*. The syllable *tha* (/ða/ in phonetic notation) is another potential fused response. Participants entered their responses by clicking on one of the response options with a computer mouse. If none of the syllables matched their perceptions, they were instructed to select "other" and describe in writing what they heard. All participants completed 4 practice trials before beginning the test trials. The nine stimuli were repeated four times, yielding 36 trials that were presented in random order.

RESULTS

A preliminary analysis confirmed that vocalized syllables were highly intelligible in spoken and sung conditions. Of participants who judged auditory unimodal stimuli (n = 14), the percentage of correctly identified syllables was near perfect for the in-key and out-of-key conditions (for both, M = 98.21, SE = 1.21) and somewhat lower for the spoken condition (M = 90.18, SE = 2.67) [$F(2,26) = 7.36, p < .003$].

For audio–visual conditions, participants also achieved near-perfect accuracy on congruent trials (M = 98.36, SE = 1.24).[1] For incongruent trials, fused responses (e.g., /da/ or /ða/) occurred on 87.50%, 87.50%, and 83.03% of trials in the spoken, in-key, and out-of-key conditions, respectively. A repeated measures ANOVA revealed no significant differences between these means ($F < 1$). As displayed in Figure 2, these percentages are significantly higher than the percentage of fused responses expected on the basis of chance (20% in a five-alternative task) for the spoken [$t(27) = 12.92, p < .001$], the in-key [$t(27) = 14.29, p < .001$], and the out-of-key [$t(27) = 10.75, p < .001$] conditions. (Note that our estimate of chance performance is conservative, because the category of "other" includes many possible percepts.) In view of high levels of /ba/ responses in the unimodal audio condition (M = 87.88, SD = 12.15), the findings illustrate clear and robust audio–visual integration across spoken and sung conditions.

To better understand the perceptual consequences of our manipulations, we next analyzed the percentage of correct (/ba/) and fused (/da/ and /ða/) responses following incongruent audio–visual presentations. First, the percentage of correct /ba/ responses was lowest for the spoken (M = 6.26, SD = 21.21), intermediate for the in-key (M = 10.71, SD = 23.98), and highest for the out-of-key (M = 12.50, SD = 25.90) condition [$F(2,54) = 2.98, p = .06$]. However, post hoc comparisons using Bonferroni correction yielded no significant differences between pairs of conditions ($p > .05$). Second, there was no significant difference in the percentage of fused /da/ responses in the spoken (M = 13.34, SD = 24.98), in-key (M = 25.00, SD = 34.02), and out-of-key (M = 28.57, SD = 32.43) conditions ($F < 1$). Third, the percentage of (fused) /ða/ responses was highest for the spoken (M = 74.10, SD = 34.34), intermediate for the in-key (M = 62.50, SD = 38.78), and lowest for the out-of-key (M = 52.67, SD = 35.57) condition [$F(2,54) = 3.493, p < .05$]. However, post hoc comparisons revealed no significant difference between pairs of conditions ($p > .05$).

DISCUSSION

Although striking dissociations can occur in the abilities to speak and sing (Schlaug, Marchina, & Norton, 2008; Straube, Schulz, Geipel, Mentzel, & Miltner, 2008;

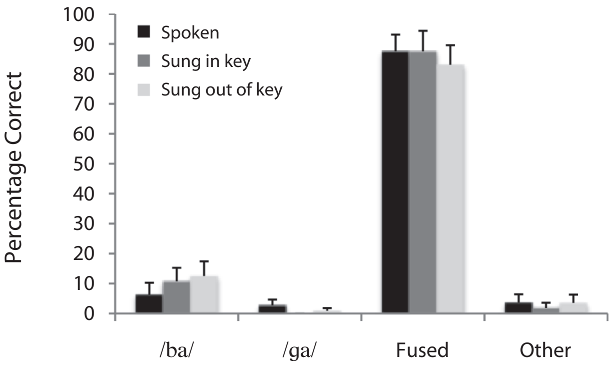

Figure 2. Percentage of syllables chosen in response to the incongruent audio–visual conditions. Standard error bars are shown.

Zipse, Norton, Marchina, & Schlaug, 2009), our results suggest that processes involved in integrating audio and visual signals of linguistic information are not domain specific. We observed a preponderance of fused responses for both spoken and sung syllables, and the rate of fused responses for sung syllables closely paralleled that for spoken stimuli. The findings illustrate, for the first time, that the McGurk effect can be observed in sung stimuli, and they imply that speech and song activate a common underlying mechanism for audio–visual integration.

The percentage of fused responses was remarkably consistent in the two domains, in spite of many surface differences. Not only did sung syllables involve a melodic sequence that was consistent with the conventions of Western tonal music, differences in vocal constraints gave rise to distinct acoustic profiles for spoken and sung syllables. For example, the degree of articulation varied significantly under the two conditions, with a longer average duration of syllables in the spoken condition. Given that these acoustic differences were not associated with observed differences in the overall rate of fused responses, we infer that the acoustic properties that varied across sung and spoken conditions are not relevant to processes of audio–visual integration.

Although we observed no differences across conditions in the percentage of combined fused responses, we observed differences in the rate of reporting one of the two fused responses—/ða/. In particular, there were relatively fewer /ða/ responses in the out-of-key condition. Audio–visual speech research has also revealed differences in the proportion of /da/ versus /ða/ responses, although these differences are variable in extent and are poorly understood (Burnham & Dodd, 2004; Green, Kuhl, Meltzoff, & Stevens, 1991; Massaro & Cohen, 1983). The type of fused response selected appears to depend on a range of factors, including the speaker, vowel type, whether visual cues were live or recorded, and how auditory stimuli were synthesized. In our investigation, differences in the type of fused response could relate to any of the above factors, including how the performer approached the phonemic information in the sung conditions. Alternatively, they may relate to attentional capture by an unexpected pitch contour.

Cross-modal experiences are ubiquitous. In most circumstances, the auditory and visual signals are consistent with one another. This redundancy is especially valuable if signals arising from one modality are weak or unreliable (Massaro, 2004), and the influence of one modality tends to increase when it is perceived to be more reliable (Alais & Burr, 2004). Although some differences in the type of fused response were observed between conditions, the data illustrate remarkable similarity in the overall occurrence of fused responses in spoken and sung contexts.

AUTHOR NOTE

Data collection and manuscript preparation were supported by discovery grants from the Natural Sciences and Engineering Research Council of Canada and by an Australian Research Council grant to W.F.T. We thank Michael Hall for helpful comments on earlier versions of the manuscript. We thank Dmitri Tcherbadji and Catherine Greentree for their assistance in recording and in manuscript preparation. Address correspondence concerning this article to W. F. Thompson, Department of Psychology, Macquarie University, Sydney, NSW 2109, Australia (e-mail: bill.thompson@mq.edu.au).

REFERENCES

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, **14**, 257-262. doi:10.1016/j.cub.2004.01.029

Besson, M., Faïta, F., Peretz, I., Bonnel, A.-M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, **9**, 494-498. doi:10.1111/1467-9280.00091

Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D. A., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition*, **81**, B11-B20. doi:10.1016/S0010-0277(01)00117-2

Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer (Version 4.6.29) [Computer program]. Available at www.praat.org.

Bonnel, A.-M., Faïta, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception & Psychophysics*, **63**, 1201-1213.

Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, **45**, 204-220. doi:10.1002/dev.20032

Collister, L. B., & Huron, D. (2008). Comparison of word intelligibility in spoken and sung phrases. *Empirical Musicology Review*, **3**, 109-125.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, **24**, 433-454. doi:10.1525/mp.2007.24.5.433

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, **21**, 103-113. doi:10.1177/030573569302100201

Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, K. N. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception & Psychophysics*, **50**, 524-536.

Gregg, J. W., & Scherer, R. C. (2006). Vowel intelligibility in classical singing. *Journal of Voice*, **20**, 198-210. doi:10.1016/j.jvoice.2005.01.007

Hidalgo-Barnes, M., & Massaro, D. W. (2007). Read my lips: An animated face helps communicate musical lyrics. *Psychomusicology*, **19**, 3-12.

Hollien, H., Mendes-Schwartz, A. P., & Nielsen, K. (2000). Perceptual confusions of high-pitched sung vowels. *Journal of Voice*, **14**, 287-298. doi:10.1016/S0892-1997(00)80038-7

Jesse, A., & Massaro, D. W. (2010). Seeing a singer helps comprehension of the song's lyrics. *Psychonomic Bulletin & Review*, **17**, 323-328.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

MacDonald, J., Andersen, S., & Bachmann, T. (2000). Hearing by eye: How much spatial degradation can be tolerated? *Perception*, **29**, 1155-1168. doi:10.1068/p3020

Massaro, D. W. (2004). From multisensory integration to talking heads and language learning. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of multisensory processes* (pp. 153-176). Cambridge, MA: MIT Press.

Massaro, D. W., & Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 753-771. doi:10.1037/0096-1523.9.5.753

Massaro, D. W., Cohen, M. M., & Smeele, P. M. T. (1996). Perception of asynchronous and conflicting visible and auditory speech. *Journal of the Acoustical Society of America*, **100**, 1777-1786. doi:10.1121/1.417342

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

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, **58**, 351-362.

NEELY, K. K. (1956). Effect of visual factors on the intelligibility of speech. *Journal of the Acoustical Society of America*, **28**, 1275-1277. doi:10.1121/1.1908620

OSTRY, D. J., & MUNHALL, K. G. (1994). Control of jaw orientation and position in mastication and speech. *Journal of Neurophysiology*, **71**, 1528-1545.

PARÉ, M., RICHLER, R. C., TEN HOVE, M., & MUNHALL, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception & Psychophysics*, **65**, 553-567.

PATEL, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, **6**, 674-681. doi:10.1038/nn1082

PATEL, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.

PATEL, A. D. (2009). Music and the brain: Three links to language. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 208-216). Oxford: Oxford University Press.

PERETZ, I., & COLTHEART, M. (2003). Modularity of music processing. *Nature Neuroscience*, **6**, 688-691. doi:10.1038/nn1083

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

SAMS, M., MÖTTÖNEN, R., & SIHVONEN, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, **23**, 429-435. doi:10.1016/j.cogbrainres.2004.11.006

SCHLAUG, G., MARCHINA, S., & NORTON, A. (2008). From singing to speaking: Why singing may lead to recovery of expressive language function in patients with Broca's aphasia. *Music Perception*, **25**, 315-323. doi:10.1525/mp.2008.25.4.315

SCHUTZ, M., & LIPSCOMB, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception*, **36**, 888-897. doi:10.1068/p5635

SMITH, A. (1992). The control of orofacial movements in speech. *Critical Reviews in Oral Biology & Medicine*, **3**, 233-267.

STRAUBE, T., SCHULZ, A., GEIPEL, K., MENTZEL, H.-J., & MILTNER, W. H. R. (2008). Dissociation between singing and speaking in expressive aphasia: The role of song familiarity. *Neuropsychologia*, **46**, 1505-1512. doi:10.1016/j.neuropsychologia.2008.01.008

SUMBY, W. H., & POLLACK, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26**, 212-215. doi:10.1121/1.1907309

SUNDBERG, J. (1977). The acoustics of the singing voice. *Scientific American*, **3**, 82-91.

'T HART, J., COLLIER, R., & COHEN, A. (1990). *A perceptual study of intonation: An experimental–phonetic approach to speech melody*. Cambridge: Cambridge University Press.

THOMPSON, W. F., GRAHAM, P., & RUSSO, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, **156**, 203-227. doi:10.1515/semi.2005.2005.156.203

THOMPSON, W. F., RUSSO, F. A., & LIVINGSTONE, S. (2010). Facial expressions of pitch structure in music performance. *Psychonomic Bulletin & Review*, **17**, 317-322. doi:10.3758/PBR.17.3.317

THOMPSON, W. F., RUSSO, F. A., & QUINTO, L. (2008). Audio–visual integration of emotional cues in song. *Cognition & Emotion*, **22**, 1457-1470. doi:10.1080/02699930701813974

VINES, B. W., KRUMHANSL, C. L., WANDERLEY, M. M., & LEVITIN, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, **101**, 80-113. doi:10.1016/j.cognition.2005.09.003

ZIPSE, L., NORTON, A., MARCHINA, S., & SCHLAUG, G. (2009). Singing versus speaking in nonfluent aphasia. *NeuroImage*, **47**, S39-S41. doi:10.1016/S1053-8119(09)71121-8

NOTE

1. For the subset of participants who judged auditory and visual unimodal stimuli in addition to congruent audio–visual stimuli ($n = 14$), accuracy across sung and spoken conditions was significantly different for the congruent audio–visual ($M = 99.70$, $SE = 1.19$), auditory unimodal ($M = 95.53$, $SE = 0.30$), and visual unimodal ($M = 63.83$, $SE = 3.93$) conditions [$F(2,26) = 70.10$, $p < .001$].

(Manuscript received May 18, 2009;
revision accepted for publication April 5, 2010.)