Current Biology 17, 1680–1685, October 9, 2007 DOI 10.1016/j.cub.2007.08.043

Report

Auditory-Visual Crossmodal Integration in Perception of Face Gender

Eric L. Smith,[1] Marcia Grabowecky,[1] and Satoru Suzuki[1,*]
[1]Department of Psychology and Institute for Neuroscience
Northwestern University
Evanston, Illinois 60208

Summary

Whereas extensive neuroscientific and behavioral evidence has confirmed a role of auditory-visual integration in representing space [1–6], little is known about the role of auditory-visual integration in object perception. Although recent neuroimaging results suggest integrated auditory-visual object representations [7–11], substantiating behavioral evidence has been lacking. We demonstrated auditory-visual integration in the perception of face gender by using pure tones that are processed in low-level auditory brain areas and that lack the spectral components that characterize human vocalization. When androgynous faces were presented together with pure tones in the male fundamental-speaking-frequency range, faces were more likely to be judged as male, whereas when faces were presented with pure tones in the female fundamental-speaking-frequency range, they were more likely to be judged as female. Importantly, when participants were explicitly asked to attribute gender to these pure tones, their judgments were primarily based on relative pitch and were uncorrelated with the male and female fundamental-speaking-frequency ranges. This perceptual dissociation of absolute-frequency-based crossmodal-integration effects from relative-pitch-based explicit perception of the tones provides evidence for a sensory integration of auditory and visual signals in representing human gender. This integration probably develops because of concurrent neural processing of visual and auditory features of gender.

Results and Discussion

Information about locations and identities of objects is often concurrently signaled by visual and auditory modalities. With respect to perception of spatial location, auditory-visual integration is well understood in terms of both the underlying neural substrates (e.g., multimodal neurons have spatially overlapped receptive fields for both visual and auditory signals) and its behavioral relevance (e.g., spatially coincident sounds enhance visual detection) [1–6]. Because many objects make characteristic sounds (e.g., cats meow and wineglasses clink), auditory-visual integration may also occur for object perception, combining object-specific signals from both modalities. Consistent with this idea, recent neuroimaging results revealed crossmodal neural activations that were object specific [7–11]. We sought to provide complementary behavioral evidence demonstrating that object perception implicitly uses crossmodally integrated sensory signals.

To behaviorally demonstrate sensory integration, one must show that perception via one modality (e.g., vision) is substantially influenced by a concurrent signal via another modality (e.g., audition). If an object could be identified unimodally through either vision or audition, however, a presumed crossmodal integration could be the result of a semantic interaction rather than a sensory integration. For a behavioral crossmodal-integration effect to provide evidence for sensory integration, the effect must occur even when object identification is possible through only one modality.

Perception of face gender, which is a salient and integral aspect of face identification [12, 13], provides an ideal test case. Gender is closely associated with sound frequencies because human adult males and females have relatively separate distributions of the dominant vocalization frequencies known as the fundamental-speaking frequencies [14] (see Figure 2). Thus, by presenting pure tones (single-frequency tones) with frequencies in the male and female fundamental-speaking-frequency ranges, we were able to present gender-specific auditory information without the spectral components that allow conscious recognition of the signal as a human voice. Furthermore, neurophysiological and neuroanatomical evidence indicates that pure tones primarily activate low-level processing in the auditory core area [15–17] and that low-level sensory processing contributes to crossmodal interactions, potentially through the diffuse "matrix" component of thalamocortical projections [18]. We thus hypothesized that if low-level auditory-frequency processing was integrated with visual face processing, a gender ambiguous (androgynous) face should appear masculine when it is accompanied by a pure tone in the male fundamental-speaking-frequency range but that the same face should appear feminine when it is accompanied by a pure tone in the female fundamental-speaking-frequency range.

Observers saw a number of briefly presented androgynous faces, which were simultaneously accompanied by either a low or high tone (Figure 1). Observers indicated the gender of each face when the face was accompanied by one specific tone (either low or high). When a face was accompanied by the other tone, observers performed a foil task of indicating the race (Asian or Caucasian) of the face. Observers were thus instructed to use the tones as task indicators. The effect of auditory frequency was examined by assignment of different frequencies to the low and high tones for different groups of observers. The pairs of low and high tones used (100 Hz and 140 Hz [both in the male range], 120 Hz and 240 Hz [one in the male range and the other in the

*Correspondence: satoru@northwestern.edu

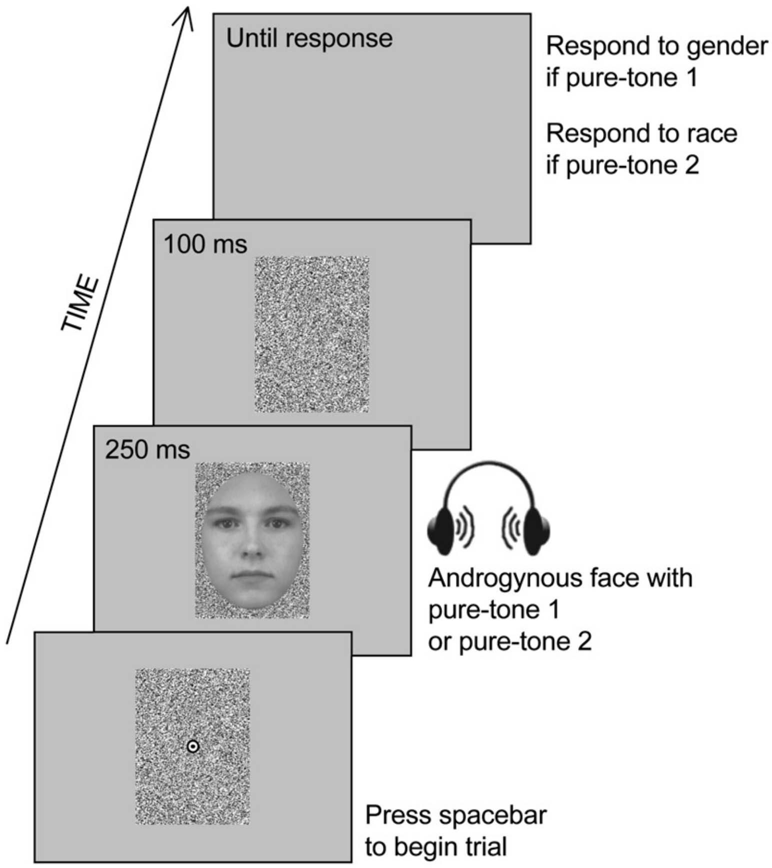

Figure 1. A Schematic Illustration of a Trial Sequence for the Crossmodal-Integration Experiment

female range], 215 Hz and 255 Hz [both in the female range], and 55 Hz and 3200 Hz [both outside of the typical speaking range]) spanned a broad range of frequencies including the male and female fundamental-speaking-frequency ranges (Figure 2). Each observer heard a specific low-high tone pair with either of the two possible tone-task combinations (the low tone indicating the gender task and the high tone indicating the race [foil] task, or vice versa). This experimental design ensured that observers paid attention to the tones but were unaware of the experimental manipulation of the tones (confirmed in postexperiment debriefing). It also allowed us to evaluate the effects of both absolute and relative frequencies of the tones.

As expected (upper panel, Figure 2), when androgynous faces were presented together with pure tones in the male fundamental-speaking-frequency range (blue region), the faces were more frequently judged as male, whereas when the faces were presented with pure tones in the female fundamental-speaking-frequency range (pink region), they were more frequently judged as female, compared to when the faces were presented with control tones that were too low (the left-most point) or too high (the right-most point) to be in the typical speaking range. To statistically confirm this frequency tuning of the crossmodal-integration effect, we conducted a two-factor ANOVA with the tone frequency (below male range, male range, female range, or above female range) and observers' gender (male or female) as the between-observer factors. The main effect of the tone frequency was significant, $F_{3,268} = 5.68$, $p < 0.001$, confirming that the effect of pure tones on the visual perception of face gender was tuned to the gender-specific fundamental-speaking-frequency ranges. Furthermore, the male (140 Hz) and female (240 Hz) peaks of the tuning function significantly deviated from the control level (55 Hz and 3200 Hz combined) ($t_{108} = 2.75$, $p < 0.05$ for the male peak, and $t_{142} = 2.82$, $p < 0.04$ for the female peak, with the p values Bonferroni corrected). In addition, female observers tended to see the faces as more feminine than did male observers (45.7% versus 35.9% female responses), $F_{1,268} = 9.93$, $p < 0.002$; this effect of observer's gender did not interact with the effect of tone frequency, $F_{3,268} = 0.05$, n.s.

To provide a concise interpretation of the data, we derived male and female crossmodal "channels" on the basis of the simple assumptions that both channels are Gaussian shaped in the log-frequency domain and that the obtained percentage of female responses was proportional to the difference in activation between the two channels. This linear dual-channel model produced a good fit to the data (the solid curve shown in the upper panel in Figure 2), and the derived tuning of the female channel closely matched the adult female fundamental-speaking-frequency range (middle panel, Figure 2). The tuning of the male channel was shifted to slightly higher frequencies relative to the adult male fundamental-speaking-frequency range. We speculate that this upward shift reflects the fact that prepuberty males have higher fundamental-speaking frequencies. The average male fundamental-speaking frequency falls

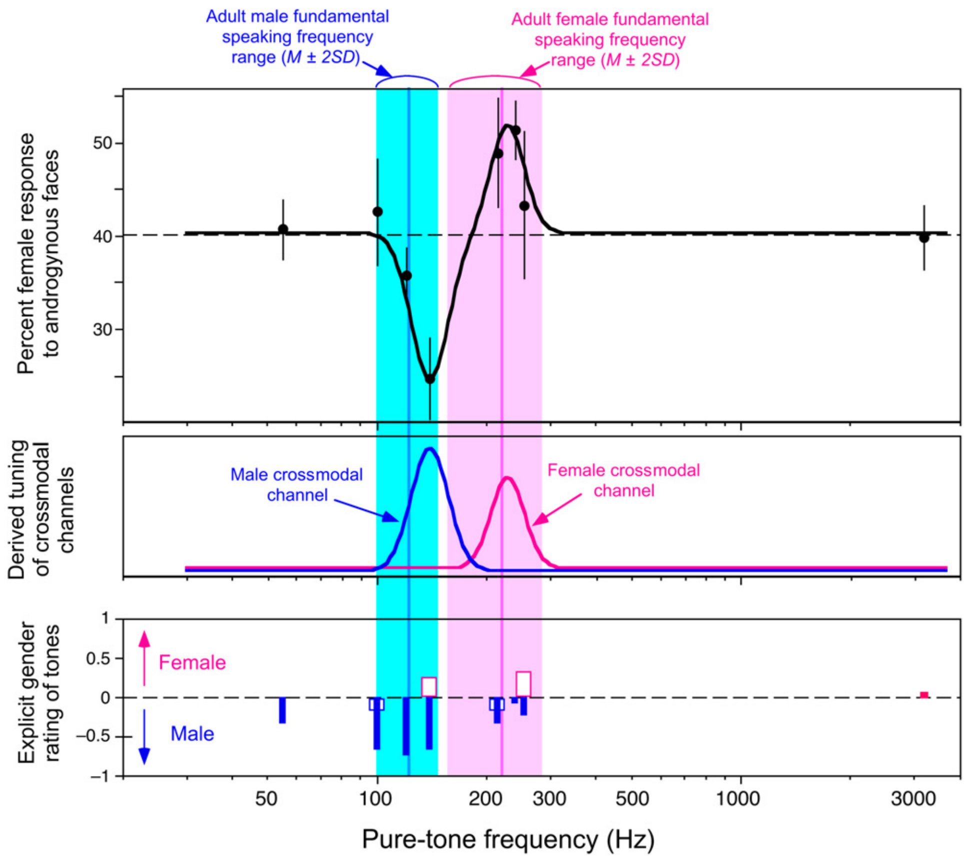

Figure 2. The Crossmodal Effect of Pure Tones on Visual Gender Perception as a Function of the Tone Frequency

The blue and pink regions indicate the normative ranges of adult male and female fundamental-speaking frequencies with the vertical lines indicating the means (based on [14]). The upper panel shows the percentage of female responses to androgynous faces as a function of the frequency of the concurrent pure tone. The error bars represent ± 1 SE. The continuous curve shows a fit based on a linear Gaussian-tuning model illustrated in the middle panel (see text for details). Note that the crossmodal-integration effect is tuned to the absolute frequencies of the male and female fundamental-speaking ranges. The lower panel shows the explicit gender ratings of the tones. Negative values represent male ratings, and positive values represent female ratings. The solid narrow bars show the gender ratings when each observer heard only one tone and rated it. Note that all tones were explicitly perceived as masculine except for the highest tone when heard first. The open wide bars show the gender ratings when the tones were heard in pairs similar to the crossmodal-integration experiment. Each observer heard the male-range pair (100 Hz and 140 Hz) or the female-range pair (215 Hz and 255 Hz) and rated the second tone of the pair in the following sequences: hearing 140 Hz then rating 100 Hz (i.e., rating the lower tone of the pair), hearing 100 Hz then rating 140 Hz (i.e., rating the higher tone of the pair), hearing 255 Hz then rating 215 Hz (i.e., rating the lower tone of the pair), or hearing 215 Hz then rating 255 Hz (i.e., rating the higher tone of the pair). Note that the lower tone of each pair was perceived as masculine (i.e., negative ratings), whereas the higher tone of each pair was perceived as feminine (i.e., positive ratings), suggesting that explicit gender impressions of the tones depend on relative frequency.

from approximately 210 Hz to the adult level of 120 Hz between ages 13 and 15 [19]. If we assume that the male crossmodal representation derives from a weighted summation of experiences with prepuberty and postpuberty males, the upward shift we obtained would be consistent with approximately 80% weight on experience with adult males and 20% weight on experience with prepuberty males.

Although pure tones do not sound like human voices, it is possible that they might nevertheless evoke a weak explicit awareness of gender. Although it is virtually impossible to demonstrate the complete absence of any gender awareness induced by the pure tones, we provide converging pieces of evidence, on the basis of postexperiment debriefing and two control experiments, that together suggest that the obtained crossmodal-integration effect reflects a sensory integration rather than a semantic interaction because of explicit auditory recognition of gender.

Our observers explicitly used the tones to decide whether to judge the gender or the race of each face, and only one tone was associated with the gender judgment for each observer. This experimental design ensured that our observers explicitly associated each tone with one of the tasks. During debriefing, none of the observers stated that they suspected any association between the tone frequency and gender. Our observers thus did not use a strategy of deliberately using the tone to judge visual gender.

To evaluate potential gender impressions of the pure tones, we explicitly asked a separate group of observers to judge the tones as to whether they fell within the typical male range, the typical female range, or outside of the speaking range (either too low to be in the male range or too high to be in the female range). Each observer rated only one pure tone because, in the crossmodal-integration experiment, only one pure tone was associated with the gender-judgment task for each observer. A different group of observers judged each of the eight tone frequencies.

To evaluate explicit judgments of the tones, we coded the male responses as −1 and the female responses as +1; we coded the outside-the-speaking-range responses as 0 to be commensurate with the crossmodal-integration effect that showed no influences from the tones outside the typical speaking range (upper panel, Figure 2). The average gender ratings were negative for all tones (see the solid narrow bars in the lower panel in Figure 2), −0.33 for 55 Hz ($t_{11} = -2.35$, n.s.), −0.67 for 100 Hz ($t_{11} = -4.69$, $p < 0.01$), −0.75 for 120 Hz ($t_{11} = -5.74$, $p < 0.005$), −0.67 for 140 Hz ($t_{11} = -4.69$, $p < 0.01$), −0.33 for 215 Hz ($t_{11} = -1.30$, n.s.), −0.08 for 240 Hz ($t_{11} = -0.32$, n.s.), and −0.25 for 255 Hz ($t_{11} = -0.90$, n.s.), except for the highest tone (3200 Hz) for which the average rating was +0.08 ($t_{11} = 1.00$, n.s.) (all p values Bonferroni corrected). Thus, all tones that produced the crossmodal-integration effect (including the tones in the female fundamental-speaking-frequency range) were explicitly judged (on average) as masculine or gender ambiguous.

We note that, although none of the tones was judged as feminine, the average rating of the tones in the female range, −0.22 (for the 215 Hz, 240 Hz, and 255 Hz tones combined), was overall less negative than the average rating of the tones in the male range, −0.69 (for the 100 Hz, 120 Hz, and 140 Hz tones combined), $t_{70} = 2.80$, $p < 0.01$, indicating that higher-frequency tones sounded less masculine. However, the pattern of tone ratings is inconsistent with the pattern of the crossmodal-integration effect in other aspects. For example, although the masculine ratings for the 100 Hz and 140 Hz tones were equally strong and both statistically significant, the 140 Hz tone produced the crossmodal-integration effect, but the 100 Hz tone did not. Furthermore, the average rating of the female-range tones was not significantly different from the average rating of the control tones (the 55 Hz [below male] and 3200 Hz [above female] tones combined), $t_{58} = 0.49$, n.s., and if anything, the average rating of the female-range tones, −0.22, was lower than the average rating of the control tones, −0.13. In contrast, the female-range tones significantly increased female responses to the androgynous faces compared to the control tones, $t_{176} = 2.55$, $p < 0.02$. Thus, overall the explicit ratings of the tones are inconsistent with the frequency tuning of the crossmodal-integration effect.

To seek further dissociation between the crossmodal-integration effect and explicit gender impressions of the tones, we examined the effect of relative frequency. The crossmodal-integration effect did not depend on relative frequency. The two male-range tones, 100 Hz and 140 Hz, the two female-range tones, 215 Hz and 255 Hz, and the two outside-the-speaking-range tones, 55 Hz and 3200 Hz, were paired so that when one tone signaled the gender task, the other tone signaled the foil (race) task. If the crossmodal-integration effect was driven by relative frequency, the proportion of female responses should have been lower for the 100 Hz tone than for the 140 Hz tone, lower for the 215 Hz tone than for the 255 Hz tone, and lower for the 55 Hz tone than for the 3200 Hz tone. None of these relative-frequency-based predictions were supported by our crossmodal-integration result (upper panel, Figure 2).

To determine whether explicit gender judgments depended on relative frequency, we conducted another tone-judgment experiment by using paired tones (with a new group of observers). Each observer heard two tones, making a race judgment for the first tone (as to whether the tone sounded Asian or Caucasian) and making a gender judgment for the second tone. In this way, the conditions for this paired-tone-judgment experiment closely matched the crossmodal-integration experiment. We used two of the tone pairs included in the crossmodal-integration experiment, one pair (100 Hz and 140 Hz) within the male fundamental-speaking-frequency range and the other pair (215 Hz and 255 Hz) within the female fundamental-speaking-frequency range. The gender ratings of the tones were scored in the same way as in the prior single-tone-judgment experiment.

The average gender ratings were, −0.17 for 100 Hz ($t_{11} = -1.48$, n.s.), +0.25 for 140 Hz ($t_{11} = +0.90$, n.s.), −0.17 for 215 Hz ($t_{11} = -0.69$, n.s.), and +0.33 ($t_{11} = +1.30$, n.s.) for 255 Hz. Although none of the ratings differed significantly from zero, the pattern of the ratings is clearly inconsistent with the pattern of the crossmodal-integration effect. Furthermore, the pattern of the ratings suggests that explicit gender ratings depend on relative frequency; that is, within each of the male-range and female-range pairs, the lower tone was rated as masculine (negative rating) and the higher tone was rated as feminine (positive rating) (see the open wide bars in the lower panel in Figure 2). This difference (the −0.17 average rating for the 100 Hz and 215 Hz tones versus the +0.29 average rating for the 140 Hz and 255 Hz tones) was statistically significant, $t_{46} = 2.02$, $p < 0.05$.

To evaluate this effect of relative frequency, we determined how the rating of each tone changed when it was rated as the second tone (after hearing a lower- or higher-frequency tone) in this experiment compared to when the same tone was rated as the first tone in the single-tone-judgment experiment. To measure directional changes in ratings, we separately coded the two categories of outside-the-speaking-range responses; we coded the below-the-speaking-range responses as −2 and the above-the-speaking-range responses as +2.

Both the 100 Hz and 215 Hz tones were rated as more masculine (or too low to be in the speaking range) when they were heard after hearing their paired higher-frequency tones (the 140 Hz and 255 Hz tones, respectively) than when they were heard as the first tone, with an average decrease in rating of 0.33 points. Similarly, both the 140 Hz and 255 Hz tones were rated as more feminine (or too high to be in the speaking range) when they were heard after hearing their paired lower-frequency tones (the 100 Hz and 215 Hz tones, respectively) than when they were heard as the first tone, with an average increase in rating of 0.92 points. In short, the tones were rated as more masculine when they were heard after hearing higher-frequency tones, whereas the tones were rated as more feminine when they were heard after hearing lower-frequency tones. This relative-frequency effect was confirmed by the significant interaction between the relative frequency (a higher or lower tone within a pair) and the presentation order (heard as the first or second tone), $F_{1,92} = 8.60$, $p < 0.005$. Thus, the crossmodal-integration effect dissociated from explicit impressions of the tones in that the former did not depend on relative frequency but the latter did.

To summarize, we generated converging evidence against the possibility that the crossmodal-integration effect we obtained might be mediated by explicit gender impressions from the pure tones. First, pure tones (which sound like beeps) do not sound like human vocalization. Second, our postexperimental interviews confirmed that none of our observers was aware of any association between the tone frequency and face gender. Third, the explicit ratings of the tones (whether each tone was heard in isolation or in the context of the tone paired in the crossmodal-integration experiment) were overall inconsistent with the frequency tuning of the crossmodal-integration effect. Fourth, the crossmodal-integration effect showed no dependence on relative frequency (despite the fact that observers attended to relative frequency on each trial to determine which of the two tasks to perform). In contrast, explicit ratings of the tones showed consistent dependence on relative frequency (despite the fact that there was no task demand for attending to relative frequency). Taken together, these pieces of evidence (even if each piece is not conclusive) support the conclusion that the crossmodal-integration effect we obtained is primarily due to sensory integration; that is, the visual face processing was fundamentally dependent on the concurrent processing of gender-consistent auditory-frequency signals.

It is noteworthy that the crossmodal-integration effect and explicit impressions of the tones dissociated with respect to their dependencies on absolute and relative frequencies. The relative-frequency dependence of the explicit gender impressions of the tones is consistent with the fact that acoustic features associated with conscious processing, for example features associated with perception of speech and music, are primarily carried by spatiotemporal changes in frequencies (rather than by absolute frequencies). It would thus be beneficial for explicit perception of acoustic signals to be primarily based on relative frequency. We have demonstrated, however, that when behaviorally significant information is carried by absolute frequency, as in gender identification, absolute-frequency information can also contribute to explicit perception by implicitly influencing another modality (vision in this case).

In conclusion, we have demonstrated that pure tones, which are processed in low-level auditory brain areas [15–17] and lack any spectral characteristics that allow auditory identification as human voices, substantially influence visual perception of face gender. The perceptual dissociation of the crossmodal effect of the tones (based on absolute frequency) from the explicit perception of the tones (based on relative frequency) provides further evidence of sensory integration. The overlap between the frequency tuning of the crossmodal-integration effect and the male and female fundamental-speaking-frequency ranges suggests that the underlying auditory-visual integration develops because of concurrent neural processing of visual gender and gender-associated auditory frequencies. Although prior research primarily focused on crossmodal integration in the perception of space [1–6], our results provide evidence of crossmodal integration in the perception of human gender, a behaviorally significant category of object perception. It is likely that crossmodal integration is ubiquitous in object perception whenever signals from multiple modalities redundantly support a behaviorally relevant perceptual classification.

Experimental Procedures

(1) Crossmodal-Integration Experiment

Observers

Two hundred and seventy-six Northwestern University undergraduate students (138 women) gave informed consent to participate for partial course credit. They all had normal or corrected-to-normal visual acuity and hearing and were tested individually in a normally lit room.

Stimuli

Auditory Stimuli. Eight pure (sinusoidal) tones were used, three tones with frequencies within the male fundamental-speaking-frequency range (100 Hz, 120 Hz, and 140 Hz), and three tones with frequencies within the female fundamental-speaking-frequency range (215 Hz, 240 Hz, and 255 Hz). The remaining pure tones had frequencies outside of the normal speaking range (55 Hz and 3200 Hz). All tones were presented with Sennheiser-HD265 headphones at approximately 76 dB SPL. Each observer heard one low and one high tone, presented one at a time in one of the following pairings: 100 Hz and 140 Hz, 215 Hz and 255 Hz, 120 Hz and 240 Hz, or 55 Hz and 3200 Hz. Observers were instructed to judge the gender of the simultaneously presented face when the face was presented with one tone (either the low or high tone) and to judge the race of the face (the foil task) when the face was presented with the other tone. Thus, each observer heard only one specific tone frequency to indicate the gender task, allowing us to manipulate the auditory frequency associated with gender judgments as a between-observer variable in order to keep observers unaware of the tone manipulation.

Observers were randomly assigned to one of the eight tones indicating the gender task (55 Hz, 100 Hz, 120 Hz, 140 Hz, 215 Hz, 240 Hz, 255 Hz, or 3200 Hz), except that more observers were assigned to the 120 Hz and 240 Hz tones because those frequencies fell near the middle of the typical male and female fundamental-speaking-frequency ranges and to the 55 Hz and 3200 Hz tones because it was important to show that frequencies outside the normal speaking range had no effect. Some minor variations in the number of observers also occurred across frequencies because of exclusion of observers who failed to follow the instructions (i.e., those who failed to perform the gender task and the foil task according to the assigned tones with 80% or greater accuracy). We thus had the following: n = 42 for 55 Hz, n = 22 for 100 Hz, n = 58 for 120 Hz, n = 18 for 140 Hz, n = 18 for 215 Hz, n = 52 for 240 Hz, n = 16 for 255 Hz, and n = 50 for 3200 Hz. We had an equal number of female and male observers for each frequency.

Visual Stimuli. Forty androgynous (20 Asian and 20 Caucasian) faces were generated by morphing between pairs of female and male faces with Avid's Elastic Reality software. Prior to generating morphs, we cropped all faces (grayscale photographs) into an elliptical shape to remove hair and jaw line, standardized them for average luminance (61 cd/m^2), size, and eye position, and placed them on a background of Gaussian noise. For each female-male face pair, 75 nodes were placed on key points of each face, and 50 frames of face-morph animation were generated on the basis of the progressive-sliding bilinear interpolation. The androgynous point for each morph was determined in a pilot experiment in which a separate group of observers (n = 35) found the point at which the gender was indistinguishable. Note that the proportion of female responses was approximately 40% (rather than 50%) in the crossmodal-integration experiment even when the tones were outside of the fundamental-speaking-frequency range and ineffective (upper panel, Figure 2). This moderate male bias is likely to be due to the fact that faces were presented briefly in this experiment (so that the initial perceptual impressions of the faces could be measured), whereas the observers in the pilot experiment viewed each face for several seconds while setting the gender-neutral points. A slight male bias in the setting could have been exaggerated under brief viewing because other simple shape features tend to be exaggerated under brief viewing conditions [20].

The androgynous faces were presented centrally on a 19″ CRT monitor at a viewing distance of 70 cm. Each face subtended 7.36° (vertical) × 4.91° (horizontal) of visual angle.

Procedure

Each trial began with a fixation screen. Upon the observer's button press, a face appeared for 250 ms concurrently with a tone, followed by a 100 ms visual Gaussian-noise mask (see Figure 1). Observers judged the gender of each face when cued by one tone (e.g., the high tone) and judged the race of the face (the foil task) when cued by the other tone (e.g., the low tone). Observers made their responses verbally in a forced-choice manner (responding either female or male when judging gender and responding either Asian or Caucasian when judging race). Observers were not time pressured, but they responded within a few seconds. Each of the 40 faces was presented twice across the two blocks of 40 trials, paired with one tone in block 1 and paired with the other tone in block 2. Trial and block orders were randomized across observers. Prior to these experimental trials, 20 practice trials were given with natural (unmorphed) female and male faces.

(2) Single-Tone-Judgment Experiment

Observers

Ninety-six Northwestern University undergraduate students (48 women) gave informed consent to participate for partial course credit or as volunteers.

Stimuli and Procedure

Each observer heard one of the eight tone frequencies used in the crossmodal-integration experiment. After hearing the tone (lasting 250 ms as in the crossmodal-integration experiment), observers rated it according to the following four classification categories: too low to be in the typical speaking range, within the typical adult male speaking range, within the typical adult female speaking range, or too high to be in the typical speaking range. Observers were not time pressured, but they completed the rating within a few seconds. Each tone frequency was judged by 12 observers (six women) in a between-observer design.

(3) Paired-Tone-Judgment Experiment

Observers

Forty-eight Northwestern University undergraduate students (24 women) gave informed consent to participate as volunteers.

Stimuli and Procedure

Each observer heard one of the following two pairs of tones: 100 Hz and 140 Hz or 215 Hz and 255 Hz (each tone lasting 250 ms as in the crossmodal-integration experiment). Observers judged the racial impression of the first tone (Asian or Caucasian) and then rated the gender impression of the second tone (by using the same four classification categories as in the single-tone-judgment experiment). There were four conditions: (1) a race judgment on the 140 Hz tone and then a gender rating of the 100 Hz tone, (2) a race judgment on the 100 Hz tone and then a gender rating of the 140 Hz tone, (3) a race judgment on the 255 Hz tone and then a gender rating of the 215 Hz tone, and (4) a race judgment on the 215 Hz tone and then a gender rating of the 255 Hz tone. Twelve observers (six women) were assigned to each condition in a between-observer design. Observers were not time pressured, but they rated each tone within a few seconds; the second tone was delivered several seconds after the rating of the first tone.

Acknowledgments

This research was supported by National Institutes of Health grant EY014110 and National Science Foundation grant BCS0643191.

Received: May 6, 2007
Revised: August 3, 2007
Accepted: August 14, 2007
Published online: September 6, 2007

References

1. Andersen, R.A., Snyder, L.H., Bradley, D.C., and Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annu. Rev. Neurosci. *20*, 303–330.
2. Stein, B.E. (1998). Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. Exp. Brain Res. *123*, 124–135.
3. Stein, B.E., Meredith, M.E., Huneycutt, W.S., and McDade, L.W. (1989). Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. J. Cogn. Neurosci. *1*, 12–24.
4. Driver, J., and Spence, C. (1998). Attention and the crossmodal construction of space. Trends Cogn. Sci. *2*, 254–262.
5. Frassinetti, F., Bolognini, N., and Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp. Brain Res. *147*, 332–343.
6. Bolognini, N., Frassinetti, F., Serino, A., and Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: Interaction among spatially converging audiovisual inputs. Exp. Brain Res. *160*, 273–282.
7. Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., and Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. Nat. Neurosci. *11*, 1190–1192.
8. Beauchamp, M.S., Lee, K.E., Argall, B.D., and Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. Neuron *41*, 809–823.
9. von Kriegstein, K., Kleinschmidt, A., Sterzer, P., and Giraud, A.-L. (2005). Interaction of face and voice areas during speaker recognition. J. Cogn. Neurosci. *17*, 367–376.
10. Amedi, A., von Kriegstein, K., van Atteveldt, M.N., Beauchamp, M.S., and Naumer, M.J. (2005). Functional imaging of human crossmodal identification and object recognition. Exp. Brain Res. *166*, 559–571.
11. Molholm, S., Ritter, W., Javitt, D.C., and Foxe, J.J. (2004). Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study. Cereb. Cortex *14*, 452–465.
12. O'Toole, A.J., Deffenbacher, K.A., Valentin, D., McKee, K., Huff, D., and Abdi, H. (1998). The perception of face gender: The role of stimulus structure in recognition and classification. Mem. Cognit. *26*, 146–160.
13. Ganel, T., and Goshen-Gottstein, Y. (2002). Perceptual integrity of sex and identity of faces: Further evidence for the single-route hypothesis. Journal of Experimental Psychology: Human Perception and Performance *28*, 854–867.
14. Deliyski, D. and Gress, C. D. (1997). Characteristics of Motor Speech Performance: Normative data. Presented at ASHA'97, American Speech-Language-Hearing Association, Boston, MA.
15. Wessinger, C.M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., and Rauschecker, J.P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. J. Cogn. Neurosci. *13*, 1–7.
16. Rauschecker, J.P. (1998). Cortical processing of complex sounds. Curr. Opin. Neurobiol. *8*, 516–521.
17. Kaas, J.H., Hackett, T.A., and Tramo, M.J. (1999). Auditory processing in primate cerebral cortex. Curr. Opin. Neurobiol. *9*, 164–170.
18. Schroeder, C.E., and Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. Curr. Opin. Neurobiol. *15*, 454–458.
19. Harries, M.L., Walker, J.M., Williams, D.M., Hawkins, S., and Hughes, I.A. (1997). Changes in the male voice at puberty. Arch. Dis. Child. *77*, 445–447.
20. Suzuki, S., and Cavanagh, P. (1998). A shape-contrast effect for briefly presented stimuli. Journal of Experimental Psychology: Human Perception and Performance *24*, 1315–1341.