

FACIAL ATTRACTIVENESS FACILITATES VOICE PROCESSING

Grant McGuire and Molly Babel

University of California, Santa Cruz and University of British Columbia

gmcguir1@ucsc.edu and molly.babel@ubc.ca

ABSTRACT

The role of visual information in auditory processing was explored using a cross-modal priming task where listeners identified the gender of male and female voice samples after being primed with a same gender face. The faces were previously rated as either attractive or unattractive. A facilitative effect was found for female voices, but not males.

Keywords: voice processing, attractiveness, speech perception

1. INTRODUCTION

Faces provide information for speech perception in a number of ways, including dynamic temporal information [7], visual cues to place of articulation [11, 10], and socially weighted expectations [5, 17, 6]. We focus here on the socially weighted expectations that are induced by face primes.

Many, if not all, of these face-voice associations are learned through experience as a listener/speaker in a speech community. For example, [5] demonstrate that speaker age and social class, as induced through photos of apparent speakers, affects perception of the New Zealand English NEAR/SQUARE merger-in-progress. Speakers that appear younger or as having lower SES are perceived as sounding more advanced in the merger. Listeners have expectations about how these different social groups produce these sounds and these expectations influence speech perception.

The associations between face gender and voice gender are learned after the associations between linguistic categories and face shape (e.g., rounded lips and the sound /u/) [15]. Adult listeners seem to robustly rely on gender associations and stereotypes in speech processing. Male and female faces can shift the way in which listeners interpret phonetic categories. [17] found that presenting perceivers with male and female faces modified their categorization of an /s/ - /ʃ/ continuum – male faces resulted in more /s/ responses, as the general frequency range associated with a female /ʃ/ overlaps with a potential male /s/. In a related study, [6] found similar effects of faces biasing categorization along

a vowel continuum. In both cases, the authors attribute the effects to expectations based on the sex of the talker changing how ambiguous auditory stimuli were perceived. This work has been replicated and extended by [12] who made clear these are effects of gender associations and not simply a function of normalization of apparent vocal tract size.

The conceptualization of gender is wrapped up in notions of typicality. [18] elaborates on the interconnected roles of gender and typicality in speech perception, and examines how stereotypicality in particular moderates the processing of spoken language. In a set of experiments, she asked participants to evaluate the typicality of a corpus of 20 voices (10 female) and 20 faces (10 female). Stereotypicality was determined through similarity rating and speeded gender identification tasks. Using these data, stereotypical and non-stereotypical male and female voices were selected for a shadowing task in an audio only condition and a priming condition where faces were used to prime the voices. In comparing the two conditions Strand found that the female faces had a very strong priming effect such that the stereotypical face facilitated shadowing. The male faces elicited no priming effects.

Differences in male and female voices were also found in [13]. They had listeners judge talker gender from single word productions and presented male and female faces at different stimulus onset asynchronies. Faces were presented 150ms before the audio, synchronous with the audio, or 150ms after the auditory stimulus. While their primary goal was to establish when indexical effects were active in speech processing, they found that female face effects occurred earlier than male effects, suggesting that expectations played a greater role for female voices.

[4] used time course data from mouse tracking to explore how sex atypical voices were affected by visual stimuli. In their study, sex-typical voices were manipulated to be atypical through the gender manipulation function in Praat, which manipulates formant ratios. Natural productions were made “more” sex-typical by decreasing formant ratios for male voices and increasing them for female voices, as well as doing the reverse to produce sex-atypical

stimuli. Listeners performed a gender identification task where they listened to typical and atypical voices and then used a mouse to select a male or female face. A small but significant effect was found where listeners showed some movement towards the opposite sex face when the voice was atypical. Whether this patterned differently for male and female voices was not addressed, and it may be worth noting that these voices were synthesized, compared to the natural productions of the other work cited above.

Given that gender typicality and voice processing interact, it is likely that attractiveness may affect voice processing as typicality and attractiveness are highly positively correlated for both faces [8, 9] and voices [2, 1]. Moreover it has been shown that facial attractiveness is correlated with vocal attractiveness [3, 16]. Following on the previous results, especially [18], we hypothesize that a more attractive face would facilitate the gender identification of a voice, while an unattractive face would slow such processing of the same voice. Following [1, 18], this study will also use speeded gender identification as a measure of processing fluency.

2. EXPERIMENT: GENDER CATEGORIZATION

2.1. Methods

Listeners participated in a speeded gender categorization task where they identified voices as male or female, one production at a time. Each word was paired with a same-gender attractive or unattractive face. Accuracy and reaction times were gathered and analyzed.

2.2. Stimuli

2.2.1. Faces

The four faces used in the experiment were composite faces generated from photos of White male and female trained actors from a publicly available database of 300 such composite faces [14]. The faces had been previously rated for attractiveness, in addition to other traits (e.g., dominance, trustworthiness, etc.) The faces selected for this study were the most and least attractive female and male faces.

2.2.2. Voices

A subset of 32 voices (16 male) from [1] were used; each voice produced the same 9 monosyllabic words. Based on the results of the gender fluency

task in [1] that study, the voices were ranked by overall response time for each gender and the 14 voices with the fastest and 14 with the slowest response times were removed. Thus, the remaining voices represented the centre of the distributions or, in other words, these voices were the most neutral in terms of categorization fluency for each gender.

2.3. Subjects

Listeners consisted of 34 (10 male, 24 female) native speakers of American English with no self-reported speech or hearing problems. All received course credit for participating.

2.4. Procedure

The experiment was administered in a sound-attenuated booth with three workstations using E-Prime software. Up to three subjects were run at a time. At the beginning of each trial, listeners were presented with face appearing on the screen for 16ms, followed immediately by an auditory presentation of a single word over headphones. Starting with the onset of auditory stimulus a screen was presented instructing the listener to press either button 1 or 5 on a 5 button response box. Buttons 1 or 5 corresponded to ?male? or ?female? responses, button and label were counterbalanced across subjects, but was consistent within the experiment for each listener.

Each face was paired with each voice within gender. That is, faces and voices always matched each other for gender making two conditions, Attractive and Unattractive. This resulted in 576 total trials: 16 voices x 9 words x 2 conditions x 2 genders.

2.5. Analysis and Results

Listeners were extremely accurate and fast at the task; summary statistics for proportion correct are shown in Table 1.

Table 1: Summary statistics for gender categorization task.

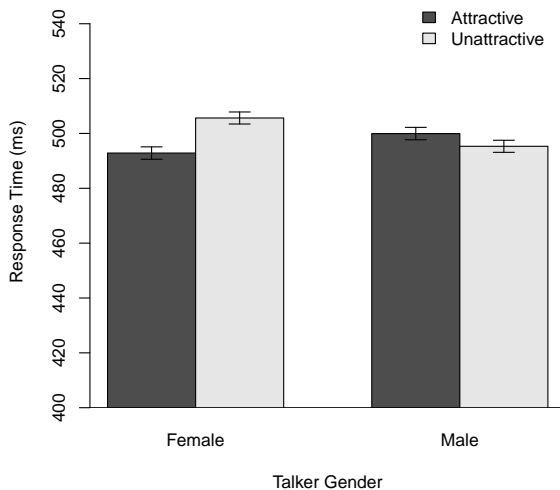
	Attractive Face	Unattractive Face
Female	0.97	0.96
Male	0.97	0.97

Reaction times for incorrect answers were removed from the data set (3%) prior to analysis. Reaction times more than 2 standard deviations from the mean were also taken out (5%). The remaining reaction times were analyzed with a mixed effect regression model having Condition (Attractive,

Unattractive) and Gender (Male, Female) as fixed effects, and random intercepts for Subject, Talker, and Word. Condition was a random slope for Subject, Talker, and Word, while Gender was a random slope for Subject. T-values were estimated by Satterthwaite approximations.

The model intercept was significant [$\beta = 495.25, SE = 13.41, z = 36.92, p < 0.001$] and there was an effect of Condition (Unattractive; [$\beta = 11.97, SE = 3.86, z = 3.10, p < 0.01$]). There was also a significant interaction between Condition (Unattractive) and Gender (Male) [$\beta = -16.83, SE = 4.56, z = -3.69, p < 0.001$]. This interaction is shown in Figure 1. Listeners were faster in identifying female voices as female when paired with the attractive female face. To confirm the effects were only present for female voices, we ran additional models identical to the initial model, but separately for male and female voices. Condition (Unattractive) was significant in the female voices model [$\beta = 11.927, SE = 4.14, t = 2.88, p < 0.01$] and was not a reliable effect in the male model.

Figure 1: Interaction between talker gender and face prime. Listeners were faster at identifying female voices as female when presented with an attractive female face. There is not difference in face type for the male voices.



3. DISCUSSION AND CONCLUSION

Listeners were faster at identifying voice gender when shown an attractive face than when shown an unattractive face. This effect was only found for female voices – male voices demonstrated no effect.

In particular the results of the current study replicate those of [18] in that an effect was only found for female voices, despite the fact that this study used more voices and more stimuli per voice. One possibility for this effect is that both studies drew faces from corpora where the male extremes were not as perceptually distant from each other as the female ones, minimizing any male effects. Another possibility is that female voices are more influenced by stereotype/attractiveness effects than male ones.

Related to the latter possibility, [1] found a positive between gender categorization fluency and gender stereotypicality only for male voices – that is the more stereotypical a male voice was rated the easier it was to identify as male. The authors reason that stereotypes about males are more grounded in experience while stereotypes about female voices are not similarly gathered from direct experience. Female voice stereotypes could be created by media exposure, something we could perhaps dub the Jessica Rabbit-effect. In their view, and assuming an exemplar theoretic model of speech perception, performance in the on-line gender fluency task is closely linked to previous, direct experience with similar voices while the off-line typicality rating task is rooted in more abstracted stereotypes. Assuming this account, female voices would be more susceptible to stereotype effects – that a typical female is attractive, in this case – and thus attractive female faces facilitate processing of female voices. Male voices, on the other hand, would not be swayed by such effects, but would be processed based on previous exemplars.

It is also possible that the asymmetry between male and female voices has psycho-acoustic roots. Female voices are psycho-acoustically less robust, and this could lead to listeners needing to develop strategies to better process female voices, whereas this compensatory strategies would not be necessary for male voices.

The underlying story behind these gender effects remains opaque, but the current study replicates an unsolved puzzle and raises questions about the role of social factors on spoken language processing. This work fits into a growing body of research which indicates that not all voices are treated equally in speech perception [19, 20].

4. REFERENCES

- [1] Babel, M., McGuire, G. in press. Perceptual fluency and judgments of vocal aesthetics and stereotypicality. *Cognitive Science*.
- [2] Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., Kawahara, H., Be-

- lin, P. 2010. Vocal attractiveness increases by averaging. *Current Biology* 20(2), 116–120.
- [3] Collins, S. A., Missing, C. 2003. Vocal and visual attractiveness are related in women. *Animal Behaviour* 65(5), 997–1004.
- [4] Freeman, J. B., Ambady, N. 2011. When two become one: Temporally dynamic integration of the face and voice. *Journal of Experimental Social Psychology* 47(1), 259–263.
- [5] Hay, J., Warren, P., Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4), 458–484.
- [6] Johnson, K., Strand, E. A., D’Imperio, M. 1999. Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics* 27(4), 359–384.
- [7] Kamachi, M., Hill, H., Lander, K., Vatikiotis-Bateson, E. 2003. Putting the face to the voice’: Matching identity across modality. *Current Biology* 13(19), 1709–1714.
- [8] Langlois, J. H., Roggman, L. A. 1990. Attractive faces are only average. *Psychological science* 1(2), 115–121.
- [9] Langlois, J. H., Roggman, L. A., Musselman, L. 1994. What is average and what is not average about attractive faces? *Psychological science* 5(4), 214–220.
- [10] McGuire, G., Babel, M. 2012. A cross-modal account for synchronic and diachronic patterns of /f/ and /θ/ in English. *Journal of Laboratory Phonology* 3, 251–272.
- [11] McGurk, H., MacDonald, J. 1976. Hearing lips and seeing voices. *Nature*.
- [12] Munson, B. 2011. The influence of actual and imputed talker gender on fricative perception, revisited (1). *The Journal of the Acoustical Society of America* 130(5), 2631–2634.
- [13] Munson, B., Solum, R. 2010. When is indexical information about speech activated? evidence from a cross-modal priming experiment. *INTERSPEECH*. Citeseer 1521–1524.
- [14] Oosterhof, N. N., Todorov, A. 2008. The functional basis of face evaluation. *Proceedings of the National Academy of Sciences* 105(32), 11087–11092.
- [15] Patterson, M. L., Werker, J. F. 2003. Two-month-old infants match phonetic information in lips and voice. *Developmental Science* 6(2), 191–196.
- [16] Saxton, T. K., Debruine, L. M., Jones, B. C., Little, A. C., Roberts, S. C. 2009. Face and voice attractiveness judgments change during adolescence. *Evolution and Human Behavior* 30(6), 398–408.
- [17] Strand, E. A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of language and social psychology* 18(1), 86–100.
- [18] Strand, E. A. 2000. *Gender stereotype effects in speech processing*. PhD thesis The Ohio State University.
- [19] Sumner, M., Kataoka, R. 2013. Effects of phonetically-cued talker variation on semantic encoding. *The Journal of the Acoustical Society of America* 134(6), EL485–EL491.
- [20] Sumner, M., Kim, S. K., King, K. B. M. 2014. The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in psychology* 4.