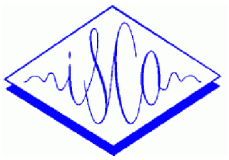

INTERSPEECH 2012
ISCA's 13th Annual Conference
Portland, OR, USA
September 9-13, 2012

# Gendered sound symbolism and masking effects in speech processing

*Molly Babel, Grant McGuire*

[1]Department of Linguistics, University of British Columbia, Vancouver, British Columbia, Canada
[2]Department of Linguistics, University of Santa Cruz, Santa Cruz, California, United States

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

## Abstract

Sound symbolism is the non-arbitrary association of sound and meaning. Experiment 1 demonstrates that sound symbolic associations facilitate the online processing of male and female voices when the target words contain the vowels /ɑ/ and /i/ for male and female voices, respectively, when listeners are engaged in a speeded gender identification task. Experiment 2 reveals that when listeners are attending to a vowel identification task, there is no effect of talker gender. This suggests support of models of perception where activation of linguistic structure masks indexical bottom-up information.

**Index Terms**: speech perception, speech processing, sound symbolism, adaptive resonance theory

10.21437/Interspeech.2012-535

## 1. Introduction

Sound symbolism is the association between sound and meaning [1]. While the existence of sound symbolism goes against Saussure's fundamental claim that the relationship between sound and meaning is arbitrary [2], there is considerable evidence that sound symbolism exists in certain limited ways [1, 3]. One such phenomenon that is relevant to this paper is the association of low back vowels like /ɑ/ with large entities and high front vowels like /i/ with small entities [4. 5]. Several scholars further link this to the sexual dimorphism found in humans where females are generally smaller than males. This plays out in names in English; the nuclear vowels in male English names are more likely to be low vowels, while the nuclear vowels of female names tend to have /i/ [6]. Moreover, a recent cross-linguistic study demonstrates that such associations between size and vowel are present in such diverse languages as Mandarin, Japanese, and Korean in addition to English [5].

While of general interest to language and cognition, there has been relatively little experimental work on sound symbolism and its relationship to speech processing (but see [7]) and the extent to which such information is used or even available at various levels of processing is unclear. In order to better understand the role of sound symbolism in speech processing we conducted two experiments using the same stimuli that require listeners to attend to different aspects of the signal: indexical voice characteristics in one, and linguistically meaningful phonemic information in the second. The goals of these experiments are twofold. The first is to examine the role of sound symbolic associations in English in the online processing of male and female voices. The second goal is to probe how participants attention to the stimuli moderates the relationship between lower-level acoustic information like talker gender, which is argued to be available before phonemic categorization [8], and comparatively higher-level linguistic information like vowel category. From our interpretation of Adaptive Resonance Theory (ART) [9, 10, 11], it is predicted that the activation of higher-level structure can interfere with the processing of low-level information. Given the previous research, we predict that sound associations between male∼/ɑ/ and female∼/i/ will surface when listeners are engaged in a speeded gender identification task (Experiment 1), but that attention to the vowel category in a speeded vowel identification task (Experiment 2) will render any sound associations between low-level information like talker gender and vowel categories opaque.

## 2. Experiment 1: Gender Identification

In the first experiment listeners were presented with the voices of sixty talkers (female = 30) producing monosyllabic words containing the vowels /i/, /ɑ/, and /u/. The task was speeded gender identification [12]; listeners were instructed to identify whether the talker that produced each word was male or female. It is predicted that this decision process will be facilitated for female voices when the target word contains the vowel /i/ and will be facilitated for male voices when the target word contains /ɑ/. It is less clear how listeners will treat /u/ considering it is both high and back. Additionally, /u/s lip rounding may make calculating vocal tract length less reliable than /i ɑ/ and therefore less susceptible to this particular sound symbolic phenomenon.

### 2.1. Methods

#### *2.1.1. Subjects*

Forty (female = 31) native speakers of North American English participated in this study. They were compen-

sated for their time with course credit.

#### 2.1.2. Stimuli

As part of a previous study [13], 30 male and 30 female voices were recorded reading 50 low frequency (from CELEX [14]) monosyllablic words with /i æ ɑ o u/ as syllable nuclei. Female talkers (mean age 24.2, range 18-57) and male talkers (mean age 24.1, range 18-47) did not differ significantly in age ($\chi$ (17) = 17.7, p = 0.4). The sixty talkers were all native speakers of American English with no reported speech or hearing problems and were compensated $10. Words were presented in 36-point font on the middle of the screen in a random order. Recordings were made directly to a computer hard drive at 44.1kHz sampling rate with a head-mounted AKG C520 microphone positioned three inches from the talkers mouth in a sound-attenuated room.

For the current experiment a subset of 9 words containing the vowels /i ɑ u/ were used; these are shown in Table 1. All tokens were normalized to have the same RMS amplitude and had silence trimmed by hand from the beginning and end of each file.

Table 1: *Words used in Experiments 1 and 2.*

| /i/ | /ɑ/ | /u/ |
|---|---|---|
| deed | cot | boot |
| key | pod | hoop |
| peel | tot | zoo |

#### 2.1.3. Procedure

Participants were seated at a computer workstation outfitted with headphones and a signal response box. Each trial consisted of a single word; the entire experiment contained 540 trials (9 words x 60 talkers). Upon the presentation of each word, participants judged whether the talker who produced the word was male or female and indicated their choice by pressing the appropriately labeled buttons on a signal-response box. The buttons were counter-balanced across participants. Participants were instructed to respond as quickly as possible while minimizing errors. Feedback on accuracy and reaction time were presented after each trial to help motivate participant performance and reduce boredom.

### 2.2. Analysis and Results

Overall, participants' accuracy was 98%. From the correct responses, reaction times more than two standard deviations from the median were removed, resulting in a loss of 5.5% of the data set. The remaining data were entered into a repeated-measures analysis of variance with reaction time as the dependent measure and vowel category (i, ɑ, u) and talker gender (male or female) as within-subject factors. There was a significant interaction between vowel and talker gender (F[2, 122] = 28.66, $p < 0.001$). This interaction is shown in Figure 1. Post-hoc Tukey tests confirmed that listeners identified male voices more quickly when the word contained an /a/ nucleus ($p < 0.001$) and that female voices were identified more quickly in /i/ environments ($p < 0.01$). There were no differences in male and female voices in the /u/ context.

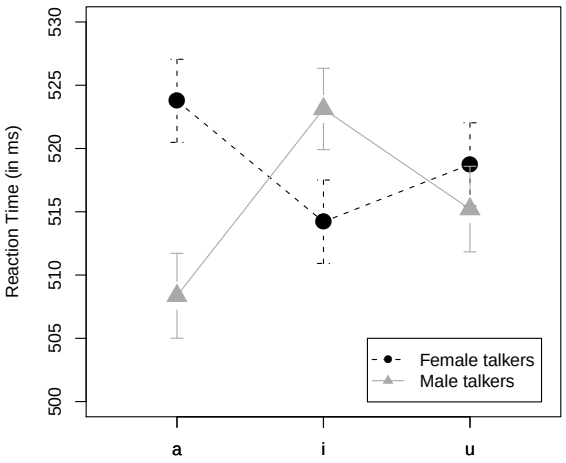


Figure 1: *Interaction between vowel category and talker gender in Experiment 1.*

### 2.3. Discussion

Experiment 1 demonstrated that male voices are easier to identify as male when the target word contains /ɑ/ and female voices are easier to identify when the word contains /i/.

Importantly, this interaction between vowel category and talker gender was found when listeners were explicitly instructed to attend to talker gender. These results suggest that linguistic knowledge is activated and interferes with processing even when not germane to the task.

## 3. Experiment 2: Vowel Identification

Experiment 1 revealed a significant interaction between talker gender and vowel such that male voices were more quickly identified in the context of /ɑ/ while female voices were more easily identified in the context of /i/. This effect was found when listeners were explicitly attending to talker gender as part of the experimental task. In the second experiment participants are presented with the same stimuli but are asked to identify the vowels in the words. Due to the masking effects of top-down in-

formation, it is predicted that by explicitly attending to a linguistic category like vowel, listeners decisions will not be affected by the gender of the talker. That is, we predict to find no interaction between vowel and talker gender in this task.

### 3.1. Methods

#### *3.1.1. Subjects*

Twenty-nine (female = 18) native speakers of North American English participated in this study. Subjects received $10 for volunteering their time.

#### *3.1.2. Stimuli*

The same stimuli as Experiment 1 was used in this task.

#### *3.1.3. Procedure*

Participants were seated at a computer workstation outfitted with headphones and a signal response box. Like Experiment 1, in Experiment 2, each trial consisted of a single word; the entire experiment contained 540 trials (9 words x 60 talkers). Upon the presentation of each word, participants judged whether the word contained the vowel /i/, /ɑ/, or /u/ by pressing the appropriately labeled buttons on the signal-response box. Buttons where labeled with both IPA symbols (/i/, /ɑ/, /u/) and orthographic labels (<ee>, <ah>, <oo>). Label assignment was counterbalanced across subjects such that the same number of participants completed one of six different label orders. Participants were instructed to respond as quickly as possible while minimizing errors. Feedback on accuracy and reaction time were presented after each trial to help motivate participant performance and reduce boredom.

### 3.2. Analysis and Results

Overall, participants were correct in 94% of their responses. Incorrect responses were removed from the data set. Reaction time outliers over two standard deviations away from the median were removed as well, which totaled 5.7% of the correct responses subset.

The remaining data were submitted to a repeated-measures analysis of variance with reaction time as the dependent measures and vowel category (i ɑ u) and talker gender (male or female) as within-subject factors. There was a single main effect of vowel ($F[2, 53] = 19.23$, $p < 0.001$). This effect is shown graphically in Figure 2. Post-hoc Tukey tests found that /i/ was identified more quickly than both /ɑ/ and /u/ ($p < 0.001$).

### 3.3. Discussion

As predicted, when participants were attending to higher-level linguistic categories such as vowel category, there was no interaction between talker gender and vowel. Instead, when processing phonemic information, inherent vowel effects dominate. The gender-specific information that interacted with vowel information in the first experiment is masked by the top-down phonemic information in this task.

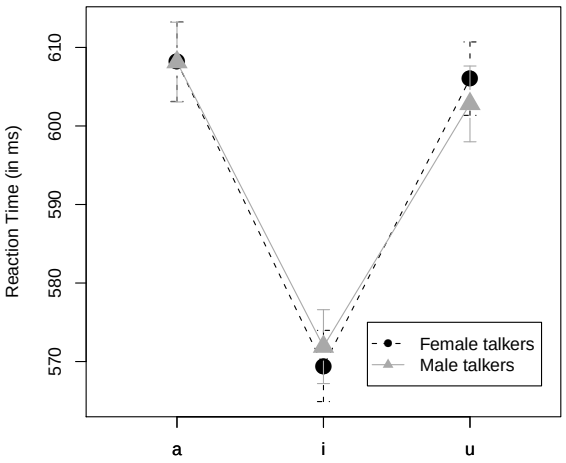


Figure 2: *Main effect of vowel in Experiment 2.*

## 4. General Discussion and Conclusion

Previous research has demonstrated an association between small stature and/or females with the high front vowel /i/ and large entities and/or males with the low back vowel /ɑ/. The vowel /u/ is often associated with round entities [15, 16], but, to our knowledge, not necessarily with males or females. Other researchers have found somewhat mixed results for this vowel, but it tends toward association with large objects [5]. In Experiment 1 we demonstrated that monosyllabic words produced by male talkers that contained /ɑ/ were identified as male voices faster than words produced by female talkers with /ɑ/. Conversely, listeners were faster at identifying female voices when the words contained the vowel /i/ than they were with male voices. The words that contained the vowel /u/ exhibited no differences for the two genders. These results suggest that sound symbolic associations facilitate speech processing. Moreover, this effect was found when listeners task involved attending to talker gender.

Experiment 2 used the exact same stimuli as Experiment 1, but listeners were instructed to identify the nuclear vowel of each word, not talker gender, as was the case for Experiment 1. When attending to the vowel category, talker gender had no effect on response time. This null result is interpreted as support for processing models

like ART, which predict that activation of top-down categories (like vowels) inhibit the influence of bottom-up information (like talker gender).

Researchers have long been aware of the phenomenon of *intrinsic pitch* or *intrinsic fundamental frequency* ($IF_0$), which is the cross-linguistic tendency for high vowels to have higher $F_0$s than low vowels. Table 2 reports mean $F_0$ for the male and female talkers for the three vowels. If the interaction of vowel and gender in the gender identification task were determined solely by $F_0$, we might, for example, expect $F_0$ of /i/ for the females to be highest, which would have facilitated a female response. The data in Table 2 illustrate, however, that /u/ has the highest $F_0$ for both female and male talkers. This set of summarized data is in line with previous work on $IF_0$ - the high vowels have higher $F_0$ than the low vowel - but they do not suggest that listeners' responses were based purely on $F_0$ cues.

Table 2: *Mean f0 in Hz (standard deviation in parentheses) for each vowel for the female and male voices.*

| | Females | Males |
|---|---|---|
| /ɑ/ | 191.473 (30) | 104.26 (16) |
| /i/ | 203.88 (27) | 110.107 (16) |
| /u/ | 213.487 (28) | 116.607 (24) |

Despite the classic view that sound and meaning are arbitrarily related in linguistic behavior, certain sound symbolic phenomena have been repeatedly demonstrated. This paper further shows that such effects, while certainly present, are mediated by the processing requirements of the task. Specifically, sound symbolic information may dominate at pre-linguistic processing, but these effects can be masked by the activation of higher-level linguistic structure. This does not necessarily mean that such information is not available to listeners, but that the requirements of the task determine whether such knowledge affects processing. These results echo other work which found that whether a sequence of two words formed a semantically cohesive compound (fire-truck vs coat-phone) interacted with listeners performance in a talker discrimination task [17]. The reverse, however, did not hold: when focused on determining whether the two word sequences formed a compound, listener responses were not influenced by whether the words were produced by the same talker. Together, these results support a model of speech recognition where linguistic information masks bottom-up information in the signal.

If these sound symbolic effects are interpreted as reflecting the availability of indexical information at different levels of speech processing, they run somewhat counter to other research. Earlier work has demonstrated phoneme judgments is affected by indexical characteristics [18, 19]. However, these sound-symbolic effects do not seem relevant to listeners during the vowel identification task and are thus discarded. This suggests that certain indexical information is not required for all speech perception tasks.

## 5. Acknowledgments

We would like to thank Steve Winters and the audiences at the 161st Meeting of the Acoustical Society of America in Seattle, WA and the Variation and Language Processing conference in Chester, UK 2011 for their feedback on this work. It was supported by funds provided by UCSC and UBC.

## 6. References

[1] Hinton, L., Nichols, J., and Ohala, J. 1994. Sound symbolic processes. In Ohala, J., Hinton, L., Nichols, J. (eds), Sound symbolism. Cambridge University Press. 1-14.

[2] Saussure, F. de. 1916. Cours de Linguistique Gnrale. C. Bally, A. Sechehaye and A. Riedlinger, eds. Paris: Payot.

[3] Sapir, E. 1929. A study in phonetic symbolism. J. of Experimental Psychology, 12, 225-239.

[4] Ohala, J. 1994. The frequency code underlies the sound-symbolic use of voice pitch. In Ohala, J., Hinton, L., Nichols, J. (eds), Sound symbolism. Cambridge University Press. 325-347.

[5] Shinohara, K., and Kawahara, S. To appear. A cross-linguistic study of sound symbolism: The images of size. Proceedings of BLS 36. Berkeley Linguistics Society.

[6] Cutler, A., McQueen, J.,and Robinson, K. 1990. Elizabeth and John: sound patterns of mens and womens names. J. of Linguistics, 26, 471-482.

[7] Westbury, C. 2005. Implicit sound symbolism in lexical access: Evidence from an interference task. Brain and Language, 93, 10-19.

[8] Goldinger, S. D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. J. of Experimental Psychology 22, 116-1183.

[9] Grossberg, S. 1980. How does a brain build cognitive code? Psychological Review, 87, 1-51.

[10] Grossberg, S. 2003. Resonant neural dynamics of speech perception. J. of Phonetics, 31, 423-445.

[11] Grossberg, S. 1999. The link between brain learning, attention, and consciousness. Consciousness and Cognition, 8, 1-44.

[12] Strand, E. 1999. Gender stereotype effects in speech processing. Ph.D. Dissertation, The Ohio State University.

[13] Babel, M. 2009. Phonetic and Social Selectivity in Speech Accommodation. Ph.D. Dissertation, UC Berkeley.

[14] Baayen, R.H., Piepenbrock, R., and van Rijn, H. 1993. The CELEX lexical database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania.

[15] Köhler, W. 1929. Gestalt Psychology. New York: Liveright.

[16] Maurer, D., Pathman, T., and Mondloch, C. J. 2006. The shape of boubas: sound-shape correspondences in toddlers and adults. Developmental Science, 9:3, 316-322.

[17] Babel, M.,and Narayan, C. 2010. Episodes and abstraction: When listeners attend to talker-specific detail. Proceedings of the Thirteenth Australasian International Conference on Speech Science and Technology, Melbourne, AU.

[18] Mullennix, J. and Pisoni, D. (1990) Stimulus variability and processing dependences in speech perception. Perception & Psychophysics, 47 (4), 379-390.

[19] Johnson, K., Strand, E., and D'Imperio, M. 1999. Auditory-visual integration of talker gender in vowel perception. J. of Phonetics, 27, 359-384.