SOUND SYMBOLISM OF GENDER IN CANTONESE FIRST NAMES

Kristen Wing Yan Wong & Yoonjung Kang

University of Toronto Scarborough & University of Toronto
wykristen.wong@mail.utoronto.ca & yoonjung.kang@utoronto.ca

ABSTRACT

Recent studies found that sound symbolism (nonarbitrary connection between sound and meaning) is more pervasive in human language than previously thought. Studies on English first names uncovered sound symbolic connection between phonological patterns and gender but comparable studies in languages other than English are limited. This study examines the sound symbolic tendencies in Cantonese first names. We built an annotated corpus of 288 commonly used disyllabic first names and examined the effect of 12 phonological characteristics (segmental, syllabic, and tonal) in predicting gender preference. Segmental effects generally conform to the predicted direction of preference (male names are more likely to contain obstruents or back/round/low vowels, or end in superheavy/closed syllables than female names) but tonal predictors (tone heights and contours) tend not to show a consistent enough effect. Also, the rhymes of the second syllable tend to play a significant role. Implications of the findings are discussed.

Keywords: Cantonese, sound symbolism, tones, gender, names

1. INTRODUCTION

The arbitrariness of the sign [9], the notion that the form or sound of a word does not resemble its meaning, is considered to be one of the major design features that sets human language apart from animal communication [14]. Studies found, however, that there are cross-linguistic tendencies to associate certain phonemes or natural classes to certain concepts, especially to concepts that have an opposing counterpart [13], and that sound symbolism is more pervasive in human language than previously thought [10, 24]. The so-called bouba-kiki (or maluma-takete) effect [18], replicated across multiple contexts [21], is a well-known example that systematic bias exists to associate certain sounds (e.g., bouba vs. kiki) with certain physical attributes (e.g., round vs. angular).

Many studies also examined sound symbolism in English personal names and found a systematic correlation between sound patterns and gender [5, 8, 29, 32, 34, 36, 37, 38]. These studies found that female names tend to be longer (in terms of syllable number); sonorants are reported to be more common in both beginnings and endings of female names, while male names have more obstruents and stop endings; and female names tend to end in an open syllable and prefer front vowels such as /i/, while male names end in a closed syllable with preference for back rounded vowels such as /ɔ/. Some of these attested properties may have universal sensory underpinnings while others may be accidental language-specific patterns with no prediction for generalizability across languages ("iconic" vs. "systematic" sound symbolism) [10]. However, linguistically-informed studies of sound-gender connections in personal names are limited outside of English (but see [35] on French names and [23] on Urdu names), and many questions remain unexplored as to the universality, the iconicity, and the psychological reality of sound-gender connections in first names.

The current study contributes toward filling this research gap by examining phonological tendencies in first names in Hong Kong Cantonese, a language genetically and typologically distinct from English. Results are compared and contrasted with previous findings from English. Furthermore, special emphasis is placed on tone since there are no similar studies conducted on tonal languages.

2. BACKGROUND

Cantonese is a Chinese language that adopts a monosyllabic tonal system, where each syllable is a morpheme and tones are phonemic. Gender is not marked grammatically in Cantonese. The Cantonese writing system, like other Chinese languages is logographic or morphographic, where each syllabic character corresponds to a morpheme [11]; Cantonese has many homophones that are identical in pronunciation but distinct in meaning [20]. This may lead one to expect that in Cantonese, and in Chinese languages more generally, the choice of male and female names relies heavily on the semantic content of the characters and the role of phonology may be minimal, if any. For example, the characters "文" and "雯" have the same pronunciation [mɐn21], but distinct meanings, 'literary' vs. 'figured clouds', and

they are predominantly used for male and female names, respectively. Given this background, Cantonese provides a more stringent test of sound symbolism in name gender, and if tendencies similar to English are also found in Cantonese, they will provide particularly strong evidence for universal patterns of sound symbolism in name gender.

In our study, we examined the different segmental, syllabic, and tonal characteristics of Cantonese names and their bias toward the male or female gender. Cantonese first names usually consist of two syllables; the current study has taken into account the possible role of syllable position in names based on the first author's native speaker intuition that the ending of a first name seems to provide more information about gender and based on the observation that the ending of a name provides more important gender cues in English [34].

3. METHOD

The name corpus used in the current study consisted of 144 disyllabic female names and 144 disyllabic male names collected from a Hong Kong blog post (https://henrycow.blogspot.com/2006/05/blog-post_12.html) that lists the 3927 most common first names generated from a survey of 3,000 Hong Kong residences in the year of 2006. Gender is not indicated in this name list; the first author, who is a native speaker, classified the list into male and female names and this classification was verified by two other native Cantonese speakers with 100% agreement. Names that can be used as either male or female names were excluded. The Cantonese names were converted to IPA transcriptions using the online conversion tool (https://toolbox.lotusfa.com/ipa/cantonese/index.html) and the transcriptions were verified against the Chinese Character Database developed by the Research Centre for the Humanities Computing of the Chinese University of Hong Kong (http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/). The names were then coded for 12 phonological characteristics, expected to correlated with gender based on previous literature, covering segmental, syllabic, and tonal features and their effects on name gender were examined. In the next section, we elaborate on how each phonological characteristic is coded and discuss the predictions based on previous literature.

3.1. Vowels

Table 1 summarizes the inventory of Cantonese monophthongal vowels and their phonological classifications adopted in our analysis. The factors examined include **vowel height** (high vs. low), **backness** (front vs. back), and **rounding** (rounded vs. unrounded). Cantonese also has a number of falling diphthongs (/ej, ɐj, a:j, u:j, ɔ:j, i:w, ɛ:w, ɐw, a:w, ow, øɥ/) and they are classified in the same way as monophthongs in our analysis, according to the nucleus.

We predict that front and unrounded vowels are favored in female names. This prediction is based on the cross-linguistic observation that sounds that are associated with high acoustic frequency, such as front unrounded vowels, high tone, and acute consonants, are used in words denoting smallness [16, 25, 26, 27, 28, 29]. High front unrounded vowels like /i/ have the highest F2 (or F2-F1 difference) while back rounded low vowels like /ɔ/ have the lowest F2 (or F2-F1). Note that rounding lowers F2. Moreover, front vowels are often associated with smallness across languages [25, 33] and front vowels, especially high front /i/, are often associated with females in sound symbolism [8, 22, 39]. This frequency-size correlation ("the frequency code") is argued to be part of human's genetic make-up due to sex-based differences in human vocal anatomy [26, 27, 28]. The universal basis for the vowel height effect is less clear and so is the evidence for it. It is, however, notable that [15] found that high vowels are correlated with smaller size in Japanese Pokémon names.

Table 1: Vowel inventory of Cantonese ([2, 4, 41])

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
High	i: ɪ e	y: ø		u: ʊ o
Low	ɛ:	œ:	ɐ a:	ɔ:

3.2. Consonants

Table 2 summarizes the inventory of Cantonese consonants and their phonological classifications adopted in our analysis. The factors examined include **sonorancy** (sonorant vs. obstruent) and **place of articulation** (grave vs. acute), in onset and coda position. We predict that obstruents and grave consonants favor male names. This is due to the fact that obstruents are associated with heaviness and largeness as well as male names [16, 25, 32, 33], and grave consonants (i.e., labial and velar places) are associated with lower frequency noise and formant transitions [28].

Table 2: Consonant inventory of Cantonese ([4, 41])

	grave	acute	grave	
	labial	coronal	dorsal	glottal
Obstruents	p p^h f	t t^h ts ts^h s	k k^h k^w k^{hw}	h
Sonorants	m w	n l j	ŋ	

3.3. Syllable structure

Cantonese lacks light syllables, where rhymes consist of a short vowel only; Cantonese syllables either contain a long vowel or end with a coda [4]. In

English, female names are more likely to end in an open syllable than male names and the final vowel in female names tends to be a schwa or a high front vowel [34, 37, 38]. Extending this observation, we examine how the composition of the rhyme affects name gender preference in Cantonese. We categorize Cantonese syllables into the following three types.

(1) Cantonese syllable types
   a. Open syllable: long vowel (**V:**)
   b. Closed syllable: short vowel + coda (**VC**)
   c. Superheavy syllable: long vowel + coda (**V:C**).

An offglide of a diphthong is analyzed as a coda following [2]. Note that it is ambiguous whether the English pattern is due to a preference for a light syllable or an open syllable in female names. We did two comparisons for the Cantonese data to test both possibilities; **open vs. closed syllables** (a vs. b-c) and **heavy vs. superheavy syllables** (a-b vs. c). Based on the English pattern, we predict that if there is any effect, closed syllables and superheavy syllables will favor male names but we note that it is not clear if and why these tendencies may be cross-linguistically available.

3.4. Tones

Table 3 summarizes the tonal inventory of Cantonese and their classification. **Tone height** divides the tones into upper and lower (or high vs. low) pitch ranges and **Tone contour** distinguishes level, rising, and falling tones. Lower frequencies, or low tones, are associated with the concept of being large and dominant while high tones tend to be associated with smallness. This correlation is expected following the hypothesis of the frequency-code [19, 26, 27, 28]. Again, equating smallness with femaleness, we expect that higher tones favor female names and lower tones favor male names. In fact, a previous study on Hong Kong Cantonese ideophones has observed that low tones are associated with the concept of largeness and masculinity while high tones are often used to express concepts related to smallness [19]. So, there are language-internal grounds for tone-gender connections as well.

In addition to pitch height, pitch contour is known to convey "social" messages; rising F0 is associated with "deference, politeness, submission, [and] lack of confidence", traits more commonly linked to females, while falling F0 is associated with "assertiveness, authority, aggression, confidence, [and] threat", more male-leaning traits [28]. Therefore, we predict that rising tones are more likely to be used in female names and falling tones are more likely to be used in male names. In addition to examining the distribution of lexical tones in each syllable, we also coded **the F0 contour across the two syllables**, as level (high-high or low-low), rising (low-high) or falling (high-low). While contour is coded as a three-way contrast, the key comparison is between rising and falling contours.

Table 3: Tone inventory of Hong Kong Cantonese (IPA tone levels are in parentheses) ([3, 4, 41])

Tone		Height	Contour
1	high-level (55/5)	high	level
2	high-rising (35)		rising
3	mid-level (33/3)		level
4	low-falling (21)	low	falling
5	low-rising (13)		rising
6	low-level (22/2)		level

3.5. Statistical Analyses

The effectiveness of predictors is examined using four different logistic regression models. (1) univariate logistic regression; (2) multivariate stepwise logistic regression using forward selection; (3) univariate mixed-effects logistic regression; (4) multivariate forward stepwise mixed-effects logistic regression. Statistical analyses were conducted in R [17]. The *glm()* function was used for regular logistic regression and the *glmer()* function of the *lme4* package [1] was used for the mixed-effects models. For the mixed-effects models, characters, separate for each of the two syllable positions, were added as random intercepts. This was done in order to avoid a few highly frequent characters skewing the results as certain characters are highly frequent in names and have a strong gender bias. For example, "偉" occurs 23 times in our database and all 23 names that contain the character are male names. In all four analyses, the dependent variable was the choice between male (=0) and female (=1) names and the predictors were one or more of the phonological predictors. All predictors were simple-coded with the male favoring levels held as the reference levels. Forward selection regression was carried out by the *step()* function.

4. RESULTS AND DISCUSSION

Figure 1 provides a descriptive summary of the gender distribution by the predictor conditions. For all predictors except for the last one (tone contour across syllables), the distribution of names is calculated separately for the first and the second syllable. For each comparison, the grey bar represents the condition that is predicted to favor male names (for example, "low" vowel) and the black bar represents the opposing condition that is predicted to favor female names ("high" vowel). The y-axis plots the proportion of female names. Overall, in most conditions, the black bars are higher than the corresponding grey bars, which means that the phonological conditions that prefer female names indeed have a higher proportion of female names.

Figure 1: Gender distribution of Cantonese names by phonological conditions (grey: male-favoring, black: female-favoring) and by syllable position

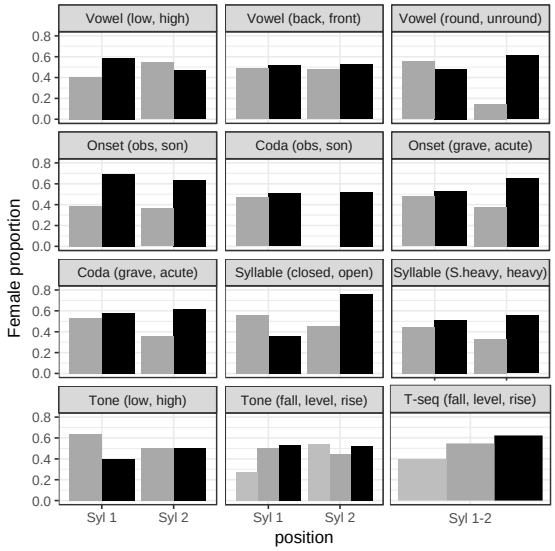

Table 4 summarizes the statistical test results. The blank cells indicate that the predictors did not show any statistically significant effect in a univariate model or the predictors did not get included in the forward selection regression in multivariate models. X marks a statistically significant difference in the opposite direction of the prediction. Checkmarks indicate a statistically significant difference in the predicted direction. Checkmarks in parentheses indicate that the predictors were selected by the stepwise regression and the coefficient is in the expected direction but did not reach statistical significance in the model. While there are a few predictors that came out in the opposite from expected direction under some tests, those are minorities (n=6) and they are outnumbered by the cases of significant effects in the predicted direction (n=32). Given the small data set and the large number of predictors, this level of positive confirmation seems encouraging.

For vowel-related factors, we found a strong effect of rounding in the second syllable and height in the first syllable. Somewhat unexpectedly, the evidence for a vowel frontness effect is marginal. This may be due to the fact that unlike English, or other languages, where a vowel backness effect is attested, Cantonese front vowels include both rounded and unrounded vowels and rounded front vowels reduce the F2 contrast between front and back vowels, thus weakening the sensory underpinnings of the sound symbolism. Future studies will take into account the co-occurrence restrictions and collinearity amongst vowel backness and rounding, as well as coda consonants.

The consonantal effects of sonorancy and place are attested quite robustly. It is notable that the sonorancy effect is more pronounced in the onset of the first syllable (beginning of the name), while the place effect is more prominent in the second syllable. The syllable structure effects are also supported, mainly for the second syllable. This is in line with the findings of English, where the ending of the name is where the rhyme plays a role in gender biases.

The intuition that the second syllable plays a more important role is partially confirmed but different factors seem important in different positions. Finally, based on the hypothesis of "frequency code" [28] and the tone-meaning connections in Cantonese ideophones [19], we had predicted that tones would have a substantial effect on name gender biases. However, the only effect in the expected direction is found for the pitch contour across two syllables—rising contour significantly prefers female over male names—but this effect is significant only under regular univariate analysis. One possible reason for the insignificant results for tones may be that tones play a subordinate role in lexical and phonological processing [37, 40].

Our study contributes to the literature on sound symbolism by examining the phonology of gender in Cantonese names. The results confirm many of the predictions; future studies will probe the interaction of related factors more closely and expand the database.

Table 4: Statistical outputs of four regression models (see text) (✓: a significant effect, (✓): a marginal effect, ×: a significant effect in the opposite direction).

Factor	Position	(1)	(2)	(3)	(4)
Vowel Height	1	✓	✓	✓	
	2				
Vowel Backness	1		×	✓	
	2				
Vowel rounding	1			×	
	2	✓	✓	✓	✓
Onset sonorancy	1	✓	✓	✓	✓
	2	✓		✓	
Coda sonorancy	1			✓	✓
	2		(✓)		
Onset place	1			×	
	2	✓	✓		✓
Coda place	1			✓	
	2	✓	(✓)	✓	
Syllable openness	1	×			
	2	✓		✓	
Syllable weight	1			✓	
	2	✓	✓	✓	✓
Tone height	1	×		×	
	2				
Tone contour	1				
	2				
T. sequence	n.a.	✓			

5. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G. 2017. lme4: Linear mixed-effects

models using 'eigen' and S4. R package, version 1.1-13. https://CRAN.R-project.org/package=lme4/
[2] Barrie, M. 2003. Contrast in Cantonese Vowels. *Toronto working papers in linguistics* 20:1-19.
[3] Barrie, M. 2007. Contour tones and contrast in Chinese Languages. *East Asian Linguist* 16:337-362
[4] Bauer, R., Benedict P. 1997. *Modern Cantonese Phonology*. Berlin: Mouton de Gruyter.
[5] Cassidy, K. W., Kelly, M. H., Sharoni, L. A. J. 1999. Inferring gender from name phonology. *Journal of Experimental Psychology: General* 128, 362–381.
[6] Carlson, R., Fant. G., Granstrom, B. 1975. Two-formant models, pitch and vowel perception. In Fant, G., Tatham, M. A. A. (eds), *Auditory analysis and perception of speech*. London: Academic Press. 55-82.
[7] Cutler, A., Mehler, J., Norris, D., Segui, J. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25, 385–400.
[8] Cutler, A., McQueen, J., Robinson, K. 1990. Elizabeth and John: Sound Patterns of Men's and Women's Names. *Journal of Linguistics* 26(2), 471-482.
[9] de Saussure, F. 1916. *Course in General Linguistics*. 1916. New York, NY: McGraw-Hill.
[10] Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., Monaghan, P. 2015. Arbitrariness, iconicity and systematicity in language. *Trends in Cognitive Sciences* 19(10), 603-615.
[11] Ettner, C. 2002. In Chinese, men and women are equal-or-women and men are equal?. In Hellinger, M. and Bussman, eds. *Gender across languages: The linguistic representation of women and men*, vol *2*, Amsterdam: J. Benjamins. 29-55.
[12] Hashimoto, O. 1972. *Phonology of Cantonese*. Cambridge: Cambridge University Press.
[13] Hinton, L., Nichols, J., Ohala, J. J. 1994. *Sound symbolism*. England: Cambridge University Press.
[14] Hockett, C. 1960. The origin of speech. *Scientific American* 203(3), 88–97.
[15] Kawahara, S., Noto, A., & Kumagai, G. 2018. Sound symbolic patterns in Pokemon names. *Phonetica* 75, 219-244.
[16] Kawahara, S., Shinohara, K. 2012. A tripartite trans-modal relationship between sounds, shapes and emotions: A case of abrupt modulation. *Proc. 34th CogSci* Austin, 569–574
[17] Kazuko S., Kawahara S. 2011. A Cross- Linguistic Study of Sound Symbolism: The Images of Size. *Proc. 36th Annual Meeting of the Berkeley Linguistics Society*. CA: UC Berkeley.
[18] Köhler, W. 1929. *Gestalt Psychology. 1st ed*. New York: Liverlight Publishing Corporation.
[19] Lai, Y.P. 2015. Phonology of Disyllabic Ideophonic Suffixes in Hong Kong Cantonese. *Language and Linguistics* 16 (5), 691-729.
[20] Li, D. C. 2017. *Multilingual Hong Kong: Languages, Literacies and Identities*. Springer International Publishing.
[21] Lockwood, G., Dingemanse, M. 2015. Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in psychology* 6, 1-14.
[22] Miall, D. S. 2001. Sounds of contrast: An empirical approach to phonemic iconicity. *Poetics* 29(1), 55-70.
[23] Mohsin, N., Kang, Y. 2018. Gender Phonology of Urdu First names. A talk at *SPF2018*, University of Toronto.
[24] Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. 2014. How arbitrary is language?. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1651), 20130299.
[25] Newman, S. S. 1933. Further experiments in phonetic symbolism. *The American Journal of Psychology* 45, 53-75.
[26] Ohala, J. J. 1983. Cross-language use of pitch: an ethological view. *Phonetica* 40. 1-18.
[27] Ohala, J. J. 1984. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41. 1-16.
[28] Ohala, J. J. 1994. The frequency code underlies thesound-symbolic use of voice pitch. In: Hinton, L., Nichols, J., Ohala, J. J. (eds), 325–347. *Sound Symbolism*. Cambridge & New York: Cambridge University Press.
[29] Pitcher, B. J., Mesoudi, A., McElligott, A.G. 2013 Sexbiased sound symbolism in English-language first names. *PLoS ONE* 8(6), e65825.
[30] R Core Team. 2016. R: A language and environment for statistical computing (Version 3.3.2) [Computer program]. Vienna, Austria: R Foundation for Statistical Computing.
[31] Sapir, E. 1929. A study in phonetic symbolism. *Journal of Experimental Psychology* 12, 225– 239.
[32] Sidhu, D. M., Pexman, P. M. 2015. What's in a name? Sound symbolism and gender in first names. *PLoS ONE* 10(5), e0126809.
[33] Shinohara, K., Kawahara, S. 2013. The sound symbolic nature of Japanese maid names. *Proc. 13th JCLA*. Japan, 183-193.
[34] Slater, A., Feinman, S. 1985. Gender and the phonology of North American first names. *Sex Roles* 13(7-8), 429-440.
[35] Sullivan, L. 2018. *Phonology of gender in English and French given names*. University of Toronto MA thesis.
[36] Sutton, L. 2016. Aliens Are Just Like Us: Personal Names in The Legion of Super Heroes. *Names* 62(2), 109-119.
[37] Wong, A. W.-K., Chen, H.-C. 2009. What are effective phonological units in Cantonese spoken word planning? *Psychonomic Bulletin & Review* 16(5), 888–892.
[38] Wright, S. K., Hay, J., & Bent, T. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 43(3), 531–561.
[39] Wu, L., Klink, R. R., Guo, J. 2013. Creating gender brand personality with brand names: The effects of phonetic symbolism. *Journal of Marketing Theory and Practice* 21(3), 319-330.
[40] Yip, M. C. W. 2001. Phonological priming in Cantonese spoken-word processing. *Psychologia* 44, 223–229.
[41] Zee, E., 1991. Chinese (Hong Kong Cantonese). *Journal of the International Phonetic Association* 21(1), 46-48.