PSYCHOACOUSTIC FEATURES EXPLAIN SUBJECTIVE SIZE AND SHAPE RATINGS OF PSEUDO-WORDS

Julián Villegas[1*] **Kimi Akita**[2] **Shigeto Kawahara**[3]
[1] Department of Computer Science and Engineering, University of Aizu, Japan
[2] Department of English Linguistics, Graduate School of Humanities, Nagoya University, Japan
[3] The Keio Institute of Cultural and Linguistic Studies, Japan

ABSTRACT

A previous study observed that the association between utterances and their ascribed meaning depends, among other factors, on phonation types. Specifically, the perceived size and shape of the referents of pseudo-words varied depending on whether they were pronounced with modal, creaky, falsetto or whisper phonation. In this study, we report a re-analysis of the same shape and size subjective ratings in terms of the four psychoacoustic measures: loudness, roughness, sharpness and pitch. We found that phonation types were positively associated with psychoacoustic features—falsetto with pitch, whisper with sharpness, and creakiness with loudness and roughness. Psychoacoustic features were also associated with subjective ratings: loudness and sharpness with subjective size and pitch and roughness with subjective shape. These findings indicate that psychoacoustic features (closer to the auditory representation of the heard sounds) may be good predictors of sound symbolism judgments.

Keywords: *Sound symbolism, psychoacoustic features, subjective shape ratings, subjective siz ratings*

1. INTRODUCTION

There seem to be non-arbitrary associations between sound and meaning. These associations are usually studied at segmental or suprasegmental level [1, 2], but more recently, the effect of phonation was investigated by Akita [3]. He studied the size and shape ratings given to pseudo-words uttered in four kinds of phonation: creaky, falsetto, modal, and whisper. It was found that creaky phonation increases the likelihood of an utterance being rated as large and more pointed than when modal phonation was used. Whispered words were, on the other hand, more likely to be rated as smaller and more rounded relative to utterances produced with modal phonation. In addition, words uttered with falsetto voice were associated with roundedness but not with smallness.

Sound and meaning associations have also been studied using acoustic measures, instead of categorical voice quality classifications. Lacey et al. [4] compared subjective shape ratings of pseudo-words with ten acoustic features including amplitude envelope, spectral tilt, harmonic-to-noise ratio, etc. They found that amplitude envelope, spectral tilt, and long-term average spectrum were associated with subjective roundness and pointedness ratings. Note that "roughness" in that research is used to describe a voice quality characterized by unevenness, aligned with the use of the same term in [5]. In our research, "roughness" is used for a psychoacoustic feature defined later.

Thus, sound and meaning associations have been studied using phonetic and phonological features which relate more to the speaker side of the speech chain [6] than to the listener side. In this study, we are interested in finding whether the ratings made by listeners in [3] could be explained by psychoacoustic models. The mental representation of acoustic phenomena is non-linear; e.g., higher frequencies are better discriminated than lower frequencies, sensitivity to intensity decreases at extreme audible frequencies, etc. Psychoacoustic models predict the auditory result of acoustic changes, i.e., they predict the probable auditory outcome reported by a listener for a given

*__Corresponding author__: *julian@u-aizu.ac.jp.*

acoustic stimulus [7].

Psychoacoustic features have been proposed for speech recognition [8], and phonation classification [9, 10], but we are not aware of previous studies using psychoacoustic features to model sound symbolic judgment patterns. For this study, we focus on the following standardized features: loudness (“that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from soft to loud”), pitch (“that attribute of auditory sensation by which sounds are ordered on the scale used for melody in music”), roughness (“subjective response to the perception of rapid amplitude modulation of a sound”) [11], and sharpness (subjective response to the spectral centroid of a sound) [7]. The main acoustic correlates of these features are sound pressure level, fundamental frequency, rapid amplitude modulations, and bass/treble ratio, respectively.

The objectives of this study are to understand how phonation types and psychoacoustic features relate to each other, and to assess how the previously reported subjective ratings on size and shape may be modelled using these features. This study is important because associations between psychoacoustic features and ascribed meaning have been under-explored, and because finding such associations could offer ways to extend the study of sound symbolism to stimuli besides speech [12].

Table 1. Size subjective ratings observed in [3]. 6 is largest.

	Creaky	Falsetto	Modal	Whisper
0	2	17	14	181
1	41	99	55	168
2	97	141	119	76
3	141	94	171	32
4	126	79	85	20
5	63	50	39	11
6	22	12	8	4

Table 2. Shape subjective ratings. 0 is roundest, 6 is the most pointed.

	Creaky	Falsetto	Modal	Whisper
0	13	60	6	39
1	14	99	43	89
2	72	149	110	136
3	82	76	137	95
4	161	71	134	83
5	120	35	60	46
6	42	14	14	16

2. METHODS

We used the same recordings and subjective ratings used in [3]. The stimuli were 48 recordings of pseudo-words (VCVs) uttered by a single male speaker of Japanese using four kinds of phonation (creaky, falsetto, modal, and whisper), 12 utterances per phonation type. These stimuli were assessed in terms of size and shape by two disjoint groups of native listeners (about 40 participants each) by means of Likert scales, as summarized in Tab. 1 and Tab. 2, respectively.

Psychoacoustic features (loudness, sharpness, pitch, and roughness) within vowels of the utterances were computed every 10 ms in Matlab [13]. The uncalibrated recordings of these utterances were assumed to be of a speaker producing modal phonation at about 65 dB(SPL) at 50 cm. Other phonations were correspondingly scaled based on their mean RMS: -10.0, 3.64, and 5.37 dB(re. mean modal phonation level) for whisper, creaky, and fasetto, respectively. In addition, the recordings were assumed to be free of reverberation. Loudness, roughness, and sharpness were computed according to Zwicker's method [14]. Pitch was computed using Stevens' method [15] from traces of fundamental frequency (F_0) extracted using Kawahara's method [16]. For the latter analysis, extreme F_0 values were set depending on phonation: 50–400 Hz for modal and whisper, 25–300 Hz for creaky, and 150–500 Hz for falsetto. The resulting psychoacoustic feature values per phonation type are shown in Fig. 1.

3. RESULTS

The collected data were analyzed with Bayesian regression models using the ‘brms’ library [17] in R [18]. Psychoacoustic measures were scaled and centered, and their mean values across uttered vowels were used as explanatory variables. We did not consider vowel quality or acoustic differences between the first and second vowels in our analysis. Further, we limited our study to only main effects (no interactions).

For computing the population-level effects, we used normal priors ($\sim N(0,1)$) for all coefficients. The convergence and mixing of MCMC chains were confirmed by inspecting effective sample sizes and $\hat{\text{R}}$-values. Fur-

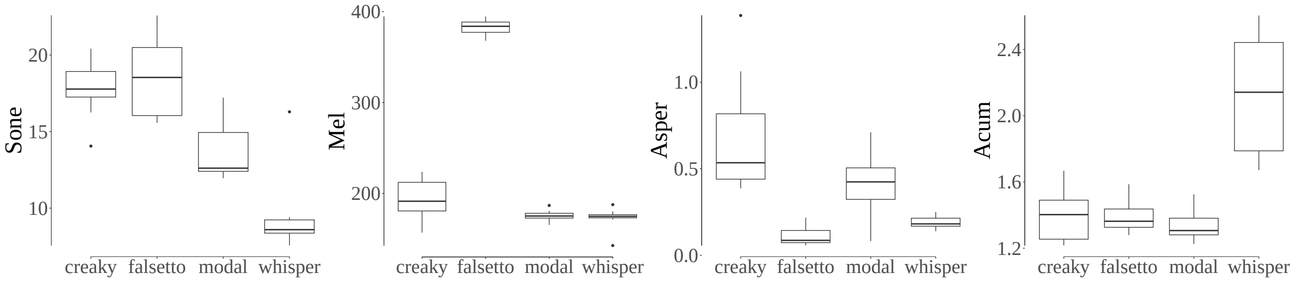

Figure 1. Psychoacoustic loudness, pitch, roughness, and sharpness for the stimuli by phonation type.

ther, the goodness of fit of the models was assessed by means of posterior predictive checks.

The effect of psychoacoustic features on phonation was analyzed using a categorical family; for their effect on subjective size and shape, a cumulative family with a logit link and flexible thresholds was used.

3.1 Phonation

The expected values from the posterior distribution (conditional effects) are presented in Fig. 2. Tab. 3 shows the regression coefficients of the model. According to this table, increasing one unit of scaled loudness increases the odds of an utterance having creaky phonation by 1.25 (in the logit scale) relative to the odds of having modal phonation when all the other psychoacoustic features were held constant. In a similar fashion, increasing scaled roughness by one unit increases the odds of creaky phonation by 0.82. Likewise, increasing one unit of scaled pitch increases the odds of an utterance having falsetto phonation by 3.04. Finally, a unit increment of scaled sharpness increases the odds of whisper phonation by 2.65.

As shown in Fig. 2, at high values of loudness, there is no difference in the probability of creaky and falsetto phonation, but the former feature is more likely than modal and whisper phonations. Falsetto phonation probability increases with pitch, as expected, whereas the probability among other phonation types was similar regardless of pitch. Similarly, the probability of creaky phonation increases with roughness while the probability of the other three phonation types is comparable regardless of roughness value, and the probability of whisper phonation increases with sharpness while the probability of other phonation types is similar regardless of sharpness value.

The association between creaky phonation and roughness has been reported before [19], but we are unaware of previous reports of whisper phonation and psychoacoustic sharpness. This outcome seems reasonable since the absence of voicing in whisper phonation increases the energy at high frequencies of the spectrum relative to the energy at low frequencies.

Table 3. Posterior means, standard deviations, and 95% credible intervals (CrI) for phonation types, using modal phonation as reference. Emboldened features here and elsewhere indicate coefficients whose CrIs do not include zero.

		μ	σ	Q.2.5	Q.97.5
Creaky	**Loudness**	1.25	0.45	0.39	2.14
	Pitch	0.60	0.57	-0.54	1.72
	Roughness	0.82	0.41	0.10	1.70
	Sharpness	0.47	0.60	-0.72	1.63
Falsetto	Loudness	0.83	0.66	-0.39	2.17
	Pitch	3.04	0.63	1.89	4.32
	Roughness	-1.07	0.79	-2.66	0.41
	Sharpness	0.04	0.80	-1.50	1.56
Whisper	Loudness	-0.45	0.53	-1.51	0.56
	Pitch	-0.85	0.81	-2.56	0.65
	Roughness	-1.08	0.75	-2.63	0.36
	Sharpness	2.65	0.66	1.44	3.98

3.2 Subjective size ratings

Rating, the dependent variable, was set as an ordered factor from 0 to 6, where 0 was associated with the smallest size and 6 to the largest. Random effects of participant and pseudo-word were included in the Bayesian modeling. Tab. 4 indicates that loudness and sharpness have credible effects on subjective size ratings. When every other psychoacoustic feature was held constant, increasing scaled loudness by one unit, increases the expected

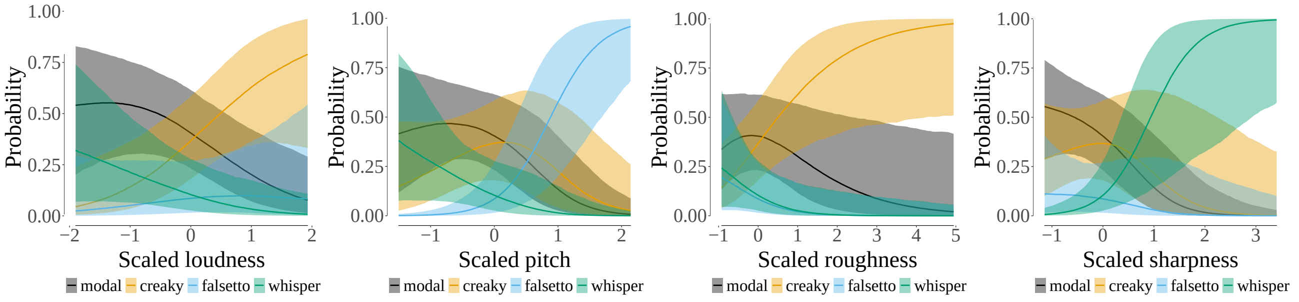

Figure 2. Phonation probability vs. four psychoacoustic features. Here and elsewhere, shaded areas around the lines correspond to 95% uncertainty intervals.

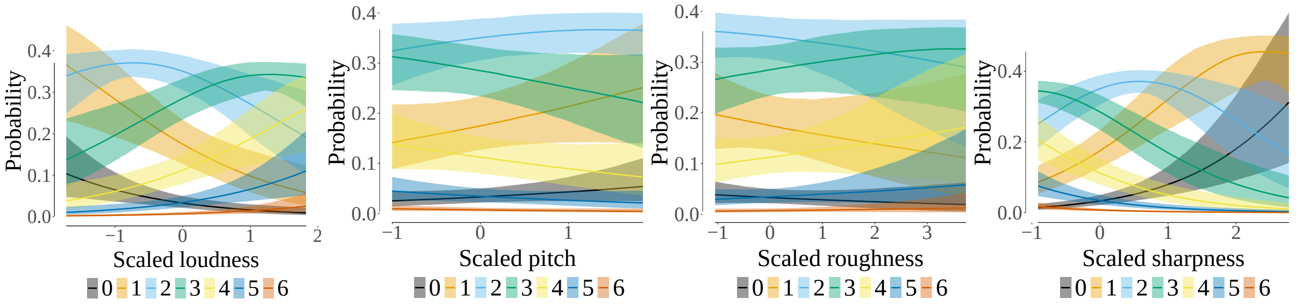

Figure 3. Subjective size vs. four psychoacoustic features. Note the different scales of the y-axes.

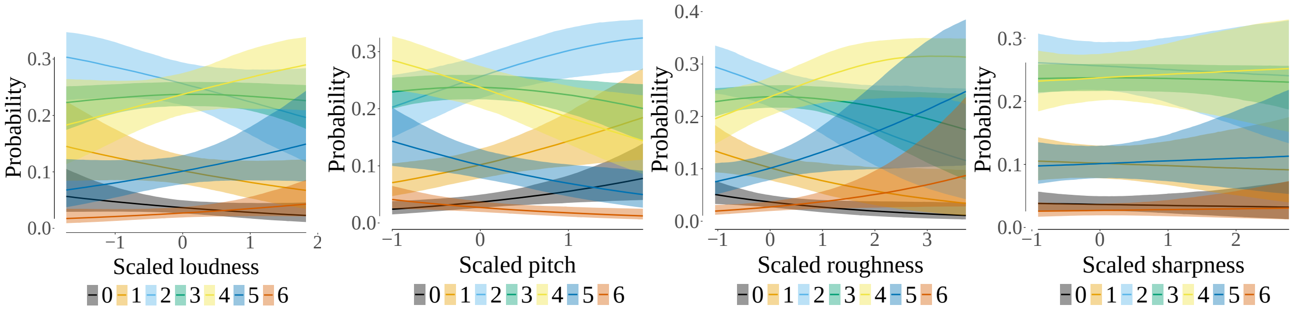

Figure 4. Subjective pointedness vs. four psychoacoustic features. Note the different scales of the y-axes.

Table 4. Posterior means, standard deviations, and 95% CrI for subjective size ratings.

	Mean	σ	Q.2.5	Q.97.5
Loudness	0.70	0.20	0.30	1.09
Pitch	−0.27	0.19	−0.63	0.10
Roughness	0.15	0.16	−0.17	0.47
Sharpness	−0.94	0.18	−1.29	−0.60

rating by 0.70 in the logit scale. In contrast, increasing one unit of the scaled sharpness decreases the expected rating by 0.94.

These results are also shown in Fig. 3. This figure shows that for loudness, extreme ratings were rather uncommon compared to ratings closer to the center of the scale. In contrast, “small” ratings (0, 1) were more likely for high values of sharpness while “large” ratings (5, 6) were not as likely regardless of sharpness value.

3.3 Subjective shape ratings

An analysis similar to that of subjective size ratings was performed for the subjective ratings of shape. In this case "rating" ranged from 0—"extremely rounded" to 6—"extremely pointed." The results of this analysis are summarized in Tab. 5 and Fig. 4. Tab. 5 indicates that while holding other features constant, increasing scaled pitch by one unit decreases the expected rating by 0.43 in the logit scale while increasing one unit of the scaled roughness increases the expected rating by 0.33.

Fig. 4 shows that extreme ratings were less likely than those closer to the center. As pitch increases, ratings below the center increase more than ratings above the center decrease. On the other hand, as roughness increases, ratings above the center increase more than ratings below the center decrease.

Table 5. Posterior means, standard deviations, and 95% CrI for subjective shape ratings.

	μ	σ	Q.2.5	Q.97.5
Loudness	0.26	0.18	−0.09	0.61
Pitch	−0.43	0.17	−0.76	−0.09
Roughness	0.33	0.15	0.02	0.63
Sharpness	0.05	0.15	−0.24	0.35

4. DISCUSSION

We found meaningful positive associations of pitch with falsetto phonation, sharpness with whisper, and loudness and roughness with creaky phonation. We also found that subjective size ratings are positively associated with loudness and negatively with sharpness, and that subjective shape ratings are negatively associated with pitch and positively associated with roughness. Recalling the results in [3], whisper phonation was more likely than modal phonation to be rated as small and rounded; creaky phonation as big and pointier, and falsetto phonation as rounded.

The present study offers a psychoacoustic explanation for the previous findings. It suggests that the apparent size of an acoustic stimulus is more clearly affected by its sharpness than by its loudness, while its pitch and roughness are rather irrelevant. The latter two features, however, credibly affect subjective shape ratings. High pitch is associated with rounded shapes and high roughness with pointed shapes.

Psychoacoustic features are universal and their models aim to predict their percepts. In addition to perception, subjective ratings are also influenced by cognition. Previous exposure to culture, language, etc. is most likely to influence these ratings. To what extent perceptual and cognitive factors affect these ratings is still unknown. Ways to shed some light on this issue are comparing the current results with ratings made by non-Japanese speakers and with those of non-verbal stimuli featuring psychoacoustic features as close as possible to the corresponding pseudo-words [20], i.e, comparing the pseudo-word ratings with ratings of psychoacoustic equivalent stimuli produced by noise, musical tones, etc. These lines of research are promising venues for future research.

One limitation of the present study is that no interactions between psychoacoustic factors were explored. Such interactions may be important for explaining the subjective ratings, but, the simpler factor-only models used here are easy to interpret and offer a good explanation of the findings. Likewise, we limit the analysis to the vowels of the pseudo-words, because phonation type more clearly manifest itself in the vowels, but the effect of consonants was found to be significant on shape ratings in [3] (albeit only in one case—velar consonants). In the future, we would perform our analysis on all the segments in the stimuli.

5. CONCLUSIONS

Psychoacoustic features can be a useful tool to model non-arbitrary associations between sound and meaning. Chiefly, subjective size ratings are affected by the loudness and sharpness of a stimulus, and subjective shape ratings are by its pitch and roughness. In addition, these features are also associated with different phonation types: loudness and roughness with creaky phonation, pitch with falsetto, and sharpness with whisper phonation. These results can be used for studying non-arbitrary associations of sound and meaning in non-verbal stimuli.

6. REFERENCES

[1] A. Nielsen and D. Rendall, "The sound of round: evaluating the sound-symbolic role of consonants in the classic takete-maluma phenomenon," *Canadian J. of Experimental Psychology*, vol. 65, no. 2, pp. 115–124, 2011. DOI: 10.1037/a0022268.

[2] B. J. Pitcher, A. Mesoudi, and A. G. McElligott, "Sex-biased sound symbolism in english-language

first names," *PLoS One*, vol. 8, no. 6, 2013. DOI: 10.1371/ journal.pone.0064825.

[3] K. Akita, "Phonation types matter in sound symbolism," *Cognitive Science*, vol. 45, no. 5, 2021. DOI: 10.1111/cogs.12982.

[4] S. Lacey, Y. Jamal, S. M. List, K. McCormick, K. Sathian, and L. C. Nygaard, "Stimulus parameters underlying sound-symbolic mapping of auditory pseudowords to visual shapes," *Cognitive Science*, vol. 44, no. 9, 2020. DOI:10.1111/cogs.12883.

[5] J. Kreiman, B. R. Gerratt, G. B. Kempster, A. Erman, and G. S. Berke, "Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research," *J. of Speech and Hearing Research*, vol. 36, pp. 21–40, 1993. DOI: 10.1044/jshr.3601.21.

[6] P. B. Denes and E. N. Pinson, *The Speech Chain: The Physics and Biology of Spoken Language*. Science/communication, New York, NY: W.H. Freeman, 2nd ed., 1993.

[7] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and models*. Berlin: Springer, 3rd ed., 2006. DOI: 10.1007/978-3-540-68888-4.

[8] E. Zwicker, E. Terhardt, and E. Paulus, "Automatic speech recognition using psychoacoustic models," *J. Acoust. Soc. Am.*, vol. 65, no. 2, pp. 487–498, 1979. DOI: 10.1121/1.382349.

[9] J. Kreiman, B. Gerratt, M. Garellek, R. Samlan, and Z. Zhang, "Toward a unified theory of voice production and perception," *Loquens*, vol. 1, no. 1, pp. 1–19, 2014. DOI: 10.3989/loquens.2014.009.

[10] J. Villegas, S. J. Lee, J. Perkins, and K. Markov, "Psychoacoustic features explain creakiness classifications made by naive and non-naive listeners," *Speech Comm.*, vol. 147, pp. 74–81, Jan. 2023. DOI: 10.1016/j.specom.2023.01.006.

[11] American National Standards Institute, "Standard Acoustical & Bioacoustical Terminology Database." ANSI/ASA S1.1 & S3.20, 2013.

[12] M. Perlman and G. Lupyan, "People can create iconic vocalizations to communicate various meanings to naïve listeners," *Scientific Reports*, vol. 8, no. 1, 2018. DOI: 10.1038/s41598-018-20961-6.

[13] Mathworks, "Matlab." Software, 2023. Version 2023a, available from `www.mathworks.com` (April 26, 2023).

[14] Int. Organization for Standardization, "Acoustics–methods for calculating loudness–Part 1: Zwicker method," *ISO 532-1:2017*, 2017.

[15] S. S. Stevens, J. Volkmann, and E. B. Newman, "A scale for the measurement of the psychological magnitude pitch," *J. Acoust. Soc. Am*, vol. 8, no. 3, pp. 185–190, 1937. DOI: 10.1121/1.1915893.

[16] H. Kawahara, Y. Agiomyrgiannakis, and H. Zen, "Using instantaneous frequency and aperiodicity detection to estimate F0 for high-quality speech synthesis," in *9th ISCA Wkshp. on Speech Synthesis*, pp. 239–246, 2016. DOI:10.21437/SSW.2016-36.

[17] P.-C. Bürkner, "Advanced Bayesian Multilevel Modeling with the R Package brms," *R Journal*, vol. 10, no. 1, pp. 395–411, 2018. Version 2.19.0.

[18] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. Version 4.2.3. Retrieved on April 26, 2023. Available from `http://www.R-project.org`.

[19] J. Villegas, K. Markov, J. Perkins, and S. J. Lee, "Prediction of creaky speech by recurrent neural networks using psychoacoustic roughness," *IEEE J. of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 355–366, 2020. DOI: 10.1109/JSTSP.2019.2949422.

[20] H. Fastl, "Neutralizing the meaning of sound for sound quality evaluations," in *Proc. Int. Congress on Acoustics*, 2001.