

TESTING THE SOUND-DRIVEN HYPOTHESIS: AN ONLINE RATING EXPERIMENT ON THE PHONAESTHETICS OF CONSTRUCTED LANGUAGES

Christine Mooshammer, Qiang Xia

Humboldt-Universität zu Berlin, Germany
christine.mooshammer@hu-berlin.de

ABSTRACT

The sound-driven hypothesis [1, 2] assumes that aesthetic judgements of languages are based on how they sound in terms of their intrinsic sound properties. For languages that are known and recognized, social factors and cultural norms play a more important role than their sound structure. In this study, the sound-driven hypothesis is tested by an online rating experiment in which 14 constructed languages, mostly from the Fantasy and Science Fiction genre, are rated regarding their pleasantness on a seven-point Likert scale. The ratings show significant correlations with several phonetic and phonological characteristics, e.g. syllable structure and sonority related measures. In order to test the dependency on the linguistic background we compared listeners from two syllable-timed tone languages, Cantonese and Mandarin Chinese, with listeners from two stress-timed intonational languages, English and German. The results indicate a stronger agreement on negative judgements whereas pleasantness depends on the L1 of the listeners.

Keywords: language attitude, phonaesthetics, typology, sonority, constructed languages

1. INTRODUCTION

Why do French and Italian sound beautiful and pleasing to most people whereas German and Arabic sound harsh and unpleasant [3, 4]? This difference in aesthetic attitude towards languages has long been discussed. Giles, Bourhis, and Davies [1] proposed the ‘inherent value hypothesis’ and the ‘imposed norm hypothesis’ to account for the causes of these language attitudes. The former hypothesis suggests that some languages have phonetic and phonological properties that are intrinsically more pleasant than others, while the latter presumes that aesthetic judgements are the result of a complex of social, cultural and political associations with the speakers of a given language. Van Bezooijen

[2] suggested ‘sound-driven hypothesis’, ‘context-driven hypothesis’ and ‘norm-driven hypothesis’ as more precise terms. Ample evidence shows that listener’s attitudes towards a language are, to some extent, biased by its imposed social prestige and cultural norms [1, 3, 5]. At least for the better-known languages that are easy to recognize, it is not straightforward to test whether the aesthetic judgement of a language is truly sound-driven. Therefore, the current study uses so-called constructed languages, i.e. languages that are invented and, in our case, used in Fantasy and Science Fiction genres. Since they are mostly unknown and unintelligible for the general public, the influence of imposed norms and subconscious associations to its speakers can be minimized in terms of language attitudes. Constructed languages (conlangs) are therefore an ideal test field to examine the sound-driven hypothesis.

Several studies have shown that phonetic and phonological characteristics contribute to aesthetic judgements of languages. Based on a sample of 16 European languages Reiterer et al. [4] found that languages with a higher sonority index [6] and a larger vocalic share are judged more positively than languages with a low sonority index and more consonants. Furthermore, a wide pitch range and frequent stressed syllables contribute to positive language perceptions [7, 8].

Certain sound-symbolic features also affect the perception of languages. For example, Crystal [9] investigated the phonological characteristics of English words that are regarded as sounding beautiful by poets and readers in newspaper polls. He found that front vowels and sonorants occur more frequently in pleasant-sounding words relative to words used in everyday conversations (see also [7] for an overview). According to Stockwell [10], guttural consonants convey a certain harshness. The production of pharyngeal and glottal sounds has been associated with negative emotions due to the similarity and the anatomical closeness with choking and coughing [11, 12]. These

sound-symbolic and phonological properties are an important design feature of conlangs [13, 14]. For example, Tolkien's Elvish languages Sindarin and Quenya [15], intended to sound beautiful, are indeed composed of a higher vocalic proportion, a higher percentage of open syllables, more sonorants, front vowels and consonants [11, 16, 17] compared to conlangs designed for evil or aggressive fictional peoples, such as Dothraki [14], Klingon [18] and Orkish [15].

The aims of this study are to test the 'sound-driven hypothesis' with languages that are unknown and unintelligible for most people, thereby minimizing the potential effect of imposed norms. Therefore, fourteen conlangs from different genres were rated in an online rating study. The second aim is to investigate whether the aesthetic judgements are affected by phonetic and phonological characteristics of the conlangs used in this study. Third, in order to test whether this hypothesis is valid for typologically diverse languages, we compare listeners from two syllable-timed tone languages, Cantonese and Mandarin Chinese [19], with listeners from two stress-timed intonation languages, English and German.

2. METHODS

2.1. Materials

In this experiment, 14 constructed languages were tested. Adûnaic, (Neo-)Khuzdul, (Neo-)Orkish, Quenya and Sindarin from the Middle-earth universe designed by J.J.R. Tolkien and featured in his work *The Lord of the Rings* [15, 20] as well as by David Salo for the film adaptation; Atlantean and Klingon by Marc Okrand [21, 18, 22]; Gaulic Vulcan [23]; Dothraki by David Peterson [14]; Na'vi by Paul Frommer [24]; Fjerdan by David Peterson and Christian Thalmann [25]; Kesh by Ursula Le Guin [26]; and the unpublished conlangs Horn and ſUiſuid by Dominique Bobeck. According to the designers' intention, Quenya and Sindarin are supposed to sound pleasant and elegant, while Dothraki, Klingon, Horn and Orkish should sound harsh, unpleasant and evil [14, 22, 15]. For each of the 14 conlangs, three sentences were recorded that were at least 10 syllables long and did not consist of any well-known buzzwords, such as *Khaleesi* in Dothraki. They were recorded by a female and a male speaker in a neutral voice without emotional involvement or sound effect at a 44100 Hz sampling rate and then downsampled to 22050 Hz for better presentation online.

2.2. Procedure

Online experiments with instruction in the four natural languages were conducted on the browser-based Percy platform [27]. Prior to the experiment participants filled out a questionnaire with information, for example about their age, gender, native language. Then participants were asked to rate the stimuli on three seven-point Likert scales: *pleasantness* (pleasant vs. unpleasant), *goodness* (good vs. evil), and *peacefulness* (peaceful vs. aggressive), solely based on their personal impression. Only the pleasantness scale will be considered here. Each stimulus could be played twice. The order of experiment stimuli and the order of three rating scales were randomized for each participant. They were offered an optional break after the 28th stimulus. The fictionality of the constructed languages was highlighted in the title of experiment 'Assess fantasy languages', so that participants were able to see it throughout the entire experiment. In total, 56 stimuli (14 conlangs x 2 sentences x 2 speakers) were rated. After finishing rating, participants had the option to listen to one additional stimulus of each conlang and guess which conlang was being played. The results are not considered here. The entire experiment took about 20 minutes. The link was distributed via email lists and social media.

2.3. Participants

After excluding non-native speakers and participants who showed a standard deviation of less than 0.5 for their ratings, the data from 22 Cantonese (YUE), 63 Chinese Mandarin (CMN), 40 English (ENG) and 91 German (GER) speakers are considered here.

2.4. Phonetic and phonological parameters

For calculating the sonority of each constructed language, we used the sonority index developed by [6]. Each speech sound group was assigned a sonority value, according to its manner of articulation, voicing and vowel height. For example, low vowels as [a, æ, ɑ] have the maximal value of 100, followed by mid back vowels [o, ɤ] with a value of 80, while voiced and voiceless stops have the minimal value of 2. The sonority index was calculated as the mean of sonority values of the phonological transcriptions of each stimulus.

Syllable boundaries were determined by the sonority profile based on the values in [6], as occurring prior to a local sonority minimum with a following rise [17]. Based on that, other indices such

as syllable rate and percentage of open syllables were derived.

With regard to phonetic properties, voiced and voiceless intervals of each recorded stimulus were manually annotated in Praat [28]. The fundamental frequency contours were calculated by using the R package wrassp [29, 30]. The f0 ranges were set to 80–500 Hz and 50–400 Hz for the female and the male speaker respectively.

3. RESULTS

3.1. Ratings

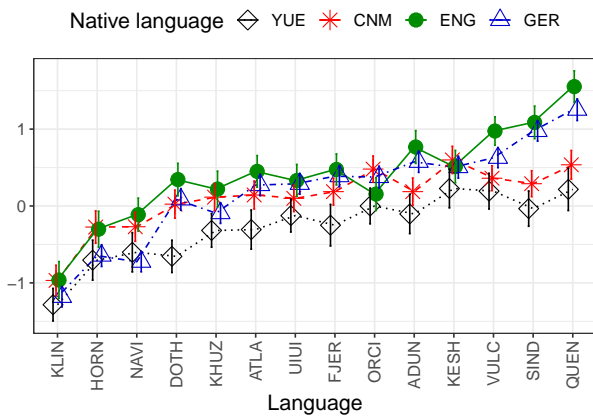


Figure 1: Mean ratings of conlangs with standard errors per listener group.

Fig. 1 shows the means and the standard errors of the pleasantness ratings for the four listener groups. As can be seen, Klingon is rated as the most unpleasant language, regardless of the L1 of the participants. The second most unpleasant conlang is Horn, except for German speaking participants who disliked Na’vi more than Horn. The most pleasant conlang for English and German speaking participants are the Elvish languages Quenya and Sindarin. Cantonese and Mandarin speaking participants chose Kesh and Quenya as most pleasant. As can be seen in Fig. 1, Cantonese and Mandarin speakers generally showed smaller positive mean ratings compared to English and German. To test whether the native language of the raters and the model speaker of the stimuli (m/f) affect the overall pleasantness ratings, a linear mixed effects model was calculated with speaker and conlang as random factors.

Fig. 2 shows the effects from the linear mixed effects model. Post hoc tests for listener group indicate that the listeners from the two tone languages Cantonese and Mandarin Chinese

rate the conlangs significantly less pleasant than listeners from the stress-timed languages English and German. There are no significant differences within the language groups. Additionally, there is a significant interaction between the factors speaker and listener group: Mandarin Chinese listeners rate stimuli from the female speaker significantly more pleasant than the male speaker.

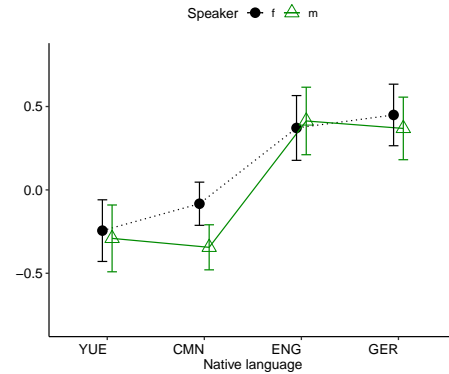


Figure 2: Effects of speaker and native language on ratings.

3.2. Relationship between ratings and phonological and phonetic characteristics

	YUE	CMN	ENG	GER
SonIndex	0.25	0.21	0.43	0.45
SonIndexCons	0.00	0.01	0.39	0.28
PctVowels	0.44	0.41	0.48	0.61
PctObstr	-0.37	-0.39	-0.73	-0.67
PctObstrOfCons	-0.23	-0.26	-0.68	-0.54
PctVoiced	0.44	0.42	0.67	0.68
PctOpenSyll	0.25	0.23	0.12	0.22
PctOpenNasSyll	0.47	0.50	0.44	0.55
PctComplOns	-0.34	-0.39	-0.52	-0.51
PctBackVOfV	-0.03	-0.07	-0.23	-0.09
PctGuttVelar	-0.40	-0.36	-0.76	-0.65
SyllRateN	0.18	0.23	0.23	0.22
F0N	-0.34	-0.37	-0.39	-0.55
IQR_F0N	-0.03	-0.01	-0.14	-0.15

Table 1: Correlation coefficients between mean ratings and phonological and phonetic characteristics of the stimuli. Coefficients with $p < 0.05$ are grey shaded, coefficients with $p < 0.001$ are additionally printed in bold.

Table 1 shows the Pearson correlation coefficients between the mean ratings and several sound-related characteristics. The ratings of all listener groups are positively affected by the percentage of vowels (PctVowels), the percentage of acoustically

measured voicing (PctVoiced) and the percentage of syllables that are open or closed by a nasal (PctOpenNasSyll). Negative correlations for all groups were found for percentage of obstruents (PctObstr), percentage of syllables with complex onsets (PctComplOns), percentage of velar and guttural sounds (PctGuttVel) and normalized F0 (F0N). Therefore, stimuli were rated as more pleasant if they contained more vowels and sonorants, longer stretches of voicing, empty or single onsets, empty or nasal codas, labial and alveolar consonants, and were spoken with a lower f0. The overall sonority (SonIndex), the sonority of the consonants (SonINdexCons) and the percentage of obstruents relative to the number of consonants in the stimulus (PctObstrOfCons) correlated only with the ratings for English and German speaking participants, not for Cantonese and Mandarin Chinese participants.

4. DISCUSSION

The first aim of this study was to test whether aesthetic judgments can be triggered based only on how the language sounds, without being recognized. The results of this rating study confirm the sound-driven hypothesis. Klingon, with the remarkable presence of uvular and guttural sounds among other things [18], was rated as the least pleasant conlang; whereas the Elvish language Quenya, having a high portion of vowels, open syllables and sonorants etc., was given a high pleasantness rating. It is noteworthy that the majority of participants was unfamiliar with the presented conlangs and did not recognize them. The linkage between aesthetic judgments and the intended impressions of Klingon speakers being aggressive and of Elves being pure and beautiful can hardly be established by the participants. Hence, biases by cultural norms and social traits can be excluded.

The second aim was to pinpoint which phonetic and phonological characteristics create the impression of pleasantness or unpleasantness. In general, stimuli with a high percentage of obstruents, complex syllable structures and velar or guttural consonants were rated less pleasant than stimuli with a large portion of voicing and vowels. This is in agreement with findings for natural languages. For example, Reiterer et al. [4] found a positive correlation between several phonological characteristics and ratings of 14 European languages. In their study, Italian was rated as beautiful which correlated with the percentage of vowels and a larger vocalic share

whereas German had the lower ratings on beauty and had also low values of sonority and vocalic share.

The third aim was to investigate whether listeners from typologically diverse languages have similar aesthetic preferences. Therefore, we compared the ratings from the two language groups, Germanic and Sinitic. The former are represented here by English and German that are stress-timed languages with an intonational prosodic system. Cantonese and Mandarin Chinese are syllables-timed languages with lexical tones and more restricted syllable structures [19]. Even though listeners from all four languages showed similar preferences, Cantonese and Mandarin Chinese listeners were more reluctant with positive judgements. As can be seen in Fig. 1, English and German listeners gave larger positive ratings than Cantonese and Mandarin Chinese listeners. One possible reason for this difference could be that none of the presented conlangs was a tonal language. Furthermore, the model speakers who produced the stimuli for this study were native speakers of German. Since the inventors of conlangs rarely describe the prosodic features of their languages, our speakers might have imposed their native German stress-timed system. This might also lead to less positive ratings from the Cantonese and Mandarin Chinese listeners. Concerning the sound-inherent phonological features, listeners from tone languages seem to be less sensitive to sonority-related measures (see Table 1) which might also account for the lower pleasantness values.

In conclusion, based on several conlangs our study confirms the sound-driven hypothesis for participants from four different languages, two tonal languages and two stress-timed languages. In a next step, we are planning to include the results from a larger variety of typologically diverse languages, such as Arabic, Russian and Italian, to further test the universality of the sound-driven hypothesis, and to examine whether phonaesthetic judgements are based on similar sets of phonological features.

5. ACKNOWLEDGEMENTS

The authors would like to express their gratitude to Christoph Draxler for his ongoing and patient support on the Percy Platform. Thanks to F. Wegmann and K. Meinhardt for recording the stimuli. We are also grateful to D. Bobeck, H. Hornecker, O. Olina, M. Walch for fruitful discussions and to the students of the class "Warum klingt Orkisch böse?" for assistance with material preparation and audio annotation.

6. REFERENCES

- [1] H. Giles, R. Bourhis, and A. Davies, “Prestige speech styles: The imposed norm and inherent value hypotheses,” in *Language and Society*. Mouton Publishers, 1979, pp. 589–596.
- [2] R. van Bezooijen, “Aesthetic evaluation of dutch,” in *Handbook of Perceptual Dialectology*, D. Long and D. R. Preston, Eds. John Benjamins Amsterdam, 2002, vol. 2, pp. 13–30.
- [3] H. Giles and N. Niedzielski, “German sounds awful, but Italian is beautiful,” in *Language Myths*, L. Bauer and P. Trudgill, Eds. Pinguin, 1998, pp. 85–93.
- [4] S. M. Reiterer, V. V. Kogan, A. Seither-Preisler, and G. Pesek, “Foreign language learning motivation: Phonetic chill or Latin lover effect? Does sound structure or social stereotyping drive FLL?” *Psychology of Learning and Motivation*, vol. 72, pp. 165–205, 2020.
- [5] A. Schüppert and G. Charlotte, “Investigating the role of language attitudes for perception abilities using reaction time,” *Dialectologia*, vol. Special Issue II, pp. 119–140, 2011.
- [6] J. G. Fought, R. L. Munroe, C. R. Fought, and E. M. Good, “Sonority and climate in a world sample of languages: Findings and prospects,” *Cross-Cultural Research*, vol. 38, no. 1, pp. 27–51, 2004.
- [7] N. H. Hilton, C. Gooskens, A. Schüppert, and C. Tang, “Is Swedish more beautiful than Danish? matched guise investigations with unknown languages,” *Nordic Journal of Linguistics*, vol. 45, no. 1, pp. 30–48, 2021.
- [8] R. van Bezooijen, “The relative importance of pronunciation, prosody, and voice quality for the attribution of social status and personality characteristics,” in *Language Attitudes in the Dutch Language Area*, R. van Hout and U. Knops, Eds. Berlin: De Gruyter Mouton, 1988, pp. 85–104.
- [9] D. Crystal, “Phonaesthetically speaking,” *English Today*, vol. 11, no. 2, pp. 8–12, 1995.
- [10] P. Stockwell, “Invented language in literature,” in *Encyclopedia of Language & Linguistics*, 2nd ed., K. Brown, Ed. Elsevier, 2006, vol. 6, pp. 3–11.
- [11] J. Podhorodecka, “Is lámatyáve a linguistic heresy? Iconicity in JRR Tolkien’s invented languages,” in *Insistent images*, O. Fischer, C. Ljungberg, and E. Tabakowska, Eds. John Benjamins Publishing, 2007, pp. 103–130.
- [12] I. Fónagy, “Why iconicity?” in *Form Miming Meaning. Iconicity in Language and Literature*, M. Nänny and O. Fischer, Eds. John Benjamins Publishing Company, 1999, pp. 3–36.
- [13] B. Beinhoff, “Why are alien languages inherently human?” *Foundation*, vol. 44, no. 122, pp. 5–19, 2015.
- [14] D. J. Peterson, *The Art of Language Invention: From Horse-Lords to Dark Elves to Sand Worms, the Words Behind World-Building*. Penguin, 2015.
- [15] J. R. R. Tolkien, *The Lord of the Rings: One volume*. Harper Collins, 2021.
- [16] H. Elsen, “Lautsymbolik in phantastischen Sprachen,” *Wirkendes Wort*, no. 1, pp. 103–119, 2019.
- [17] N. Johannesson, “Quenya, the Black Speech and the Sonority scale,” in *Proceedings of the First International Conference on J.R.R. Tolkien’s Invented Languages*, A. Stenström, Ed., pp. 14–21.
- [18] M. Okrand, *The Klingon Dictionary*, 2nd ed. Pocket Books, 1992.
- [19] P. P. K. Mok, “On the syllable-timing of Cantonese and Beijing Mandarin,” *Chinese Journal of Phonetics*, vol. 2, pp. 148–152, 2009.
- [20] J. R. R. Tolkien, *Sauron Defeated. The End of the Third Age: The History of the Lord of the Rings, part four*, ser. The History of Middle-earth, C. Tolkien, Ed. Harper Collins, 2002.
- [21] K. Shadlag. (2006) Atlantean Language Institute. [Online]. Available: https://web.archive.org/web/20140323053118/http://www.freewebs.com/keran_shadlag/
- [22] M. Okrand, *The Klingon Way: A Warrior’s Guide*. Pocket Books, 1996.
- [23] M. R. Gardner and The Vulcan Language Institute. (1980–2004) The Vulcan language. [Online]. Available: <http://surak.nu/vulcanlanguage.pdf>
- [24] S. Andrews. (2010) Paul Frommer sounds off on Avatar language. [Online]. Available: <https://news.usc.edu/32485/Paul-Frommer-Sounds-Off-on-em-Avatar-em-Language/>
- [25] C. Thalmann and D. J. Peterson. (2022) Aldjen Helt: Creating the Fjerdan language for Netflix’ Shadow and Bone. [Online]. Available: <https://sites.google.com/unito.it/i-conlangs22/book-of-abstracts>
- [26] U. K. Le Guin, *Always coming home*. London: Gollancz, 2016.
- [27] C. Draxler, “PercyConfigurator – Perception Experiments as a Service,” in *Interspeech*, 2017, pp. 823–824.
- [28] P. Boersma and D. Weenink. (2019) Praat. [Online]. Available: <http://www.praat.org/>
- [29] L. Bombien, R. Winkelmann, and M. Scheffers, *Wrassp: An R Wrapper to the ASSP Library*, 2020, R package version 0.1.9.
- [30] R. Winkelmann, J. Harrington, and K. Jänsch, “EMU-SDMS: Advanced speech database management and analysis in R,” *Computer Speech & Language*, vol. 45, pp. 392–410, 2017.