



Iconic Associations Between Vowel Acoustics and Musical Patterns, and the Musical Protolanguage Hypothesis

Gertraud Fenk-Oczlon*

University of Klagenfurt, Klagenfurt, Austria

OPEN ACCESS

Edited by:

Caicai Zhang,
The Hong Kong Polytechnic
University, Hong Kong SAR, China

Reviewed by:

Oliver Niebuhr,
University of Southern
Denmark, Denmark
Julien Meyer,
Centre National de la Recherche
Scientifique (CNRS), France

*Correspondence:

Gertraud Fenk-Oczlon
Gertraud.fenk@aau.at

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 01 March 2022

Accepted: 09 June 2022

Published: 05 July 2022

Citation:

Fenk-Oczlon G (2022) Iconic
Associations Between Vowel
Acoustics and Musical Patterns, and
the Musical Protolanguage
Hypothesis.
Front. Commun. 7:887739.
doi: 10.3389/fcomm.2022.887739

Vowels are the most musical and sonic elements of speech. Previous studies found non-arbitrary associations between vowel intrinsic pitch and musical pitch in senseless syllables. In songs containing strings of senseless syllables, vowels are connected to melodic direction in close correspondence to their *intrinsic pitch* or the frequency of the second formant F2. This paper shows that also *vowel intrinsic duration* is related to musical patterns. It is generally assumed that low vowels like [a ɔ o] have a higher intrinsic duration than high vowels like [i y u] and that there is a positive correlation between the first formant F1 and duration. Analyzing 20 traditional Alpine yodels I found that vowels with longer intrinsic duration tend to align with longer notes, whereas vowels with shorter intrinsic duration with shorter notes. This new result might shed some light on size-sound symbolism in general: Since there is a direct match between vowel intrinsic duration and the “size” of musical notes, there is no need to explain the “size” of musical notes via Ohala’s “frequency code” hypothesis. Moreover, I will argue that the iconic associations found between vowel acoustics and musical patterns support the idea of a sound-symbolic musical protolanguage. Such a protolanguage may have started with vowel syllables conveying pitch, timbre, as well as emotional, indexical, and sound-symbolic information.

Keywords: intrinsic vowel duration, size-sound symbolism, iconicity, yodels, musical notes, evolution, musical protolanguage, Ohala’s “frequency code” hypothesis

INTRODUCTION

Language and music share many commonalities, consistent with a view according to which both have a common evolutionary precursor. The hypothesized common ancestor is often referred to as “musilanguage” (Brown, 2000), “musical protolanguage” (Fitch, 2005), or “prosodic protolanguage” (Fitch, 2006). A growing number of researchers further emphasizes the idea that affective/emotional and iconic vocalizations could have played a significant role in the joint evolution of speech and music (Rousseau, 1781; Darwin, 1871; Fonagy, 1981; Levman, 1992; Scherer, 1995; Thompson et al., 2012; Perlman and Cain, 2014; Brown, 2017; Filippi and Gingras, 2018; Reybrouck and Podlipniak, 2019; Filippi, 2020).

This paper focuses on the role of vowels in the hypothetical construct “musical protolanguage.” I will briefly review some literature that has demonstrated tight relationships between vowels and music, and that has revealed the essential role of vowels in speech intelligibility

of sentences, in conveying emotional content and talker discrimination, as well as in size-sound symbolism. I then present new results showing an iconic relationship between vowel duration and musical notes in Alpine yodels. The implications for sound symbolism in general, as well as for the idea of a sound-symbolic musical protolanguage will be discussed.

The most obvious commonality between speech and music is sound, and it is the vowels that are the main carriers of sound and prosodic information in speech and singing (e.g., Fenk-Oczlon and Fenk, 2009b). Vowels are produced without obstructing the airflow from the lungs and are relatively continuous or steady-state sounds exhibiting a greater periodicity than consonants (Cutler and Mehler, 1993). According to Halle et al. (1957, p. 116) vowels can be matched easily in pitch to pure tones, whereas determinations of pitch of consonants “usually refer to the terminal stage of the second formant in the adjacent vowel.” Vowels are distinguished by their timbre, which depends on their harmonics or overtones, whereby the formants F1 and F2 are most relevant for their identification (Peterson and Barney, 1952). The main articulatory parameters responsible for vowel timbre are tongue height, front-to back position of the tongue, and lip rounding. The changes in the vowels’ resonances are audible in the case of whispering, when the vocal chords do not vibrate, or when speaking in a creaky voice (Ladefoged, 2001). Indeed, when whispering series of words like *heed, hid, head, had, hawed* one can hear the descending pitch of F2; and when speaking the series *hawed, had, head, hid, heed* in a creaky voice, the descending pitch of F1 is audible.

Timbre is clearly the primary parameter that allows for discriminating between different vowels, but vowels differ also in intrinsic pitch, intensity and duration. It is known since Meyer (1896) that, all other things being equal, high vowels such as /i/ have a higher intrinsic fundamental frequency IF0 than low vowels such as /a/. Whalen et al. (1995) could observe this effect in a sample of 31 languages and even in babbling. While the mechanism determining IF0 is still a subject of debate, there seems to be general agreement that vowel pitch depends primarily on the frequency of the second formant F2 (Marks, 1975; Traunmüller, 1986). Concerning vowel intrinsic duration it is generally assumed that low vowels have a higher intrinsic duration than high vowels like [i u y], and that there is a positive correlation between the first formant F1 and duration, i.e., the lower the vowel, the higher F1, and the higher the intrinsic duration of the vowel (House and Fairbanks, 1953; Peterson and Lehiste, 1960; Lehiste, 1970; Sol and Ohala, 2010; Toivonen et al., 2015). According to House and Fairbanks (1953) intrinsic vowel duration differences show in various types of consonant environments (voiced and voiceless stops and fricatives, nasals); for instance, when pooled across all environments the vowel /i/ has a mean duration of 0.199 s and the vowel /a/ of 0.244 s.

Evidently, vowels show all the core properties of music—timbre, intrinsic pitch, intensity and duration—and they are the most musical components of speech. Recent studies revealed tight relationships between vowels and music. For example, in Fenk-Oczlon (2017) I reported correspondences between the number of vowels and the number of pitches in musical scales across cultures: an upper limit of roughly 12 elements,

a lower limit of 2, and a frequency peak at 5 to 7 elements. The match between vowels and musical pitches shows even in specific cultures: e.g., cultures with three vowels tend to have tritonic scales. Concerning relationships between vowel acoustics and musical pitch, FÜRNISS (1991) reported associations between low vowels and the “low yodel register” and closed vowels and the “high yodel register” in the yodeling of Aka Pygmies; Fenk-Oczlon and Fenk (2009a,b) showed non-arbitrary associations between vowel intrinsic pitch and musical pitch in Alpine yodeling and in Austrian songs containing meaningless syllables. The tight bond between vowels and music is supported by experimental findings demonstrating strong interactions in the processing of vowels and melody, but not between consonants and musical information: “Vowels sing but consonants speak” (Kolinsky et al., 2009, p. 1). Similarly, Lidji et al. (2010) revealed a close processing relationship between vowels and pitch even at a pre-attentive level. Moreover, experiments by Zhang et al. (2017) demonstrated that congenital amusics not only show deficits in the perception of pitch but also in the perception of formant frequency in vowels.

Vowels and their acoustic properties are essential in many further aspects of language and speech, such as in speech intelligibility of sentences, in talker identity discrimination and in conveying emotional state, or in sound symbolism. For example, experimental studies revealed that the intelligibility of sentences was significantly better when hearing vowel-only sentences than when hearing consonant-only sentences (Cole et al., 1996; Kewley-Port et al., 2007). Vowels, unlike consonants, also provide rich indexical information about speaker identity and characteristics such as age, biological sex, origin and emotional state (Owren and Cardillo, 2006). Concerning relationships between vowels and emotional state, Rummer et al. (2014) demonstrated that subjects in a positive mood tend to invent words with /i:/, whereas when in a negative mood they tend to invent more words with /o:/.

As to sound symbolism (the non-arbitrary relation between sound and meaning), vowels are the main drivers in “size-sound symbolism” or “magnitude sound symbolism,” i.e., the association between size (large/small) and sound. In a classic study, Sapir (1929) demonstrated that participants associate meaningless words containing low and back vowels like /a/ (e.g., as in *mal*) with large concepts and meaningless words containing high and front vowels like /i/ (e.g., as in *mil*) with small concepts. Numerous experimental studies could replicate Sapir’s finding showing the postulated association between vowel quality and size (Bentley and Varon, 1933; Peña et al., 2011; Parise and Spence, 2012; Shinohara and Kawahara, 2016; Knoeferle et al., 2017; Vainio, 2021). Likewise, statistical studies in typologically diverse languages found associations between the high front vowel /i/ and the concept of small (Ultan, 1978; Haynie et al., 2014; Blasi et al., 2016; Johansson et al., 2020). Most recently, Winter and Perlman (2021) demonstrated that—in English—size adjectives clearly feature iconicity, and that the high front vowels /i/ and /I/ are associated with “small,” while the low back vowel /a/ predicts “large.” The only consonant that predicts size symbolism in their English sample was /t/. In general, consonants seem to play a rather marginal role in sound-size

associations, whereas their role in sound-shape associations as in the *maluma/takete* effect (Köhler, 1929) or the *bouba-kiki* effect (Ramachandran and Hubbard, 2001) is well-attested (but see Cuskey et al., 2017 on possible influences of orthography.)

Further cross-modal correspondences between vowels and other sensory modalities have been demonstrated between “vowels and quickness” (Jespersen, 1933), “vowels and brightness” (Marks, 1975), “vowels and spatial deixis” (Traunmüller, 1986; Johansson and Zlatev, 2013; Rabaglia et al., 2016; Vainio, 2021), “vowels and color” (Moos et al., 2014; Cuskey et al., 2019), or “vowels and taste” (Simner et al., 2010; Patak and Calvert, 2021).

Here I investigate whether there are iconic associations between the acoustic vowel property “intrinsic duration” (see above) and the length of musical notes. More specifically, I hypothesized that in songs containing meaningless syllables, syllables with low vowels like [a ɔ o] should be favored for long notes and syllables with high vowels like [i u y] for short notes.

MATERIALS AND METHODS

The singing of senseless syllables, where “the pressures of sense are relaxed to those of sound” (Butler 2015, p. 106) provides an ideal material to study relationships between vowels and musical notes. Senseless syllables are used in numerous cultures as complete or partial song texts, for example in Native American songs (Nettl, 1954), in “lilting” or “diddling,” in the singing of Scottish or Irish dance melodies, in children’s songs and jazz scat singing, or in yodeling. Here, I chose yodels for testing the hypothesized relationship between vowels and musical notes. The yodeling style, although on the whole not very frequent, can be found around the world (Grauer, 2006), for instance in Paleosiberian cultures, in the tropical forest of Africa (Pygmies), in the Kalahari Desert (Bushmen), and in the Alps (Austria, Switzerland). According to Grauer (2006) yodels are characterized across cultures by a continuous flow of sound, no embellishment, relaxed open voices, non-sense vocables, wide intervals and a polyphonic style. These characteristics also apply to traditional Alpine yodels, which are preferably polyphonic and mostly—but not necessarily—sung with frequent alternation between low and high registers (cf. Wey, 2019); they are yodeled straight without vibrato or portamento and with meaningless syllables. The yodel-syllables are predominately codaless, with rather weak or sonorant consonants in the syllabic onset, such as [jɔ, ha, hɔ, ji, ri, ho, ha]. Vowel-only syllables and codaless syllables with a liquid in the syllabic nucleus like “dl,” occur as well. The transcriptions into musical notation of the previously only orally transmitted Alpine yodels started at the beginning of the 19th century (Wey, 2019). The traditional yodels for the present study are taken from Pommer’s (1906) collection of 20 yodels. Most of the yodels of this collection are still yodeled in Austria and are well-known, so that the grapheme—phoneme correspondence of this more than 100 years old transcriptions can be checked. For instance, the grapheme “ä” is still used in Bavarian writing to denote an open “o” /ɔ/.

All 20 yodels in the collection were analyzed. I determined all relative note values in the sample: half notes (the longest note values in the sample), quarter notes, eighth notes, sixteenth notes, and thirty-second notes (the shortest notes in the sample). The notes were assigned to the respective syllables containing either high close vowels like [i u y] or low back vowels like [a ɔ o]. Furthermore, all dotted notes—the dot increases the duration of the basic note by half of its original value—were identified and matched with the particular syllables.

RESULTS

The total number of notes/syllables in the sample amounts to 1,836. The most frequent note values are eighth notes ($n = 845$), followed by quarter notes ($n = 672$), half notes ($n = 190$), sixteenth notes ($n = 95$), and thirty-second notes ($n = 34$); the number of dotted notes amounts to 348. Syllables with high vowels ($n = 1,203$) are more often used in the yodel sample than syllables with low vowels ($n = 633$); ($X^2 = 176.961, p < 0.0001$).

A detailed analysis: Eighth notes are more often aligned with high vowels (590x) than with low vowels (255x), ($X^2 = 132.811, p < 0.0001$). Quarter notes are 405 times aligned with high vowels and 267 times with low vowels ($X^2 = 28.339, p < 0.0001$). Sixteenth notes are associated with high vowels 45 times and with low vowels 50 times ($X^2 = 0.263, n.s.$). Thirty-second notes are 28 times aligned with high vowels and 6 times with low vowels ($X^2 = 14.235, p < 0.001$).

On the contrary half notes, the longest note values in the sample, are more often aligned with low vowels (135x) and less frequently associated with high vowels (55x), ($X^2 = 33.684, p < 0.0001$). This also holds for dotted notes which are 265 times associated with low vowels and only 83 times with high vowels ($X^2 = 95.184, p < 0.0001$). **Figure 1** shows an example.

DISCUSSION

Our analysis of 20 Alpine yodels demonstrates that short musical notes such as eighth notes, quarter notes and thirty-second notes tend to align with vowels with smaller intrinsic duration, whereas relative long notes such as half notes or dotted notes are associated with vowels with longer intrinsic duration. These results need to be confirmed in further studies that use an extended sample of songs containing meaningless syllables. It would also be interesting to investigate, whether in an artificial music composition game, people will tend to align vowels with longer intrinsic duration to longer notes.

Vowel Intrinsic Duration and Size-Sound Symbolism

The iconic associations between vowel intrinsic duration and length of musical notes may shed some light on size-sound symbolism in general. Although “duration” of musical notes only metaphorically corresponds to “size” of notes, our data are in line with results by Knoeferle et al. (2017) suggesting F1 and vowel duration are decisive factors in size-sound symbolism; F0 or Ohala (1984, 1994) “frequency code” hypothesis, according

7. Der Kloanschlager.

M. M. ♩ = 126. Kleinschlag*) bei Voran.

FIGURE 1 | An example of a yodeler from our sample shows that dotted and half notes tend to be linked with syllables containing the vowel *ã* /*ɔ*/ that has a longer intrinsic duration.

TABLE 1 | Examples of vowel-only sentences and vowel-only expletives in Japanese, Carinthian and in the language of the Mbendjele Pygmies.

ue o ui, o ool, ai o ou, ai ue o [worried about hunger, concealing old age, he seeks love, a love- hungry man] ooo, oooo, oo ooo [the courageous king conceals his tail when he goes out]	Japanese examples from Tsunoda (1985) cited in Bannan (2008)
a i a? Me too? “a” question particle, “i” ich (I) “a” auch (also) a e i a! Me too! “a” interjection (astonishment) :e(h) particle, “i” ich (I) “a” auch (also)	Carinthian (South Bavarian dialectal variant)
iiiiiii expletive for surprise or disgust uuuuooooo expletive to accompany a dangerous or outrageous act iiiiieeee expletive to indicate pleasure	Mbendjele Pygmies examples from Lewis (2009)

to which size-symbolism mirrors the size of the vocalizers producing either lower or higher frequencies, do not seem to play a role in their experiments on visual size judgements. Similarly, Vainio (2021) reports that F0 values did not show to be relevant

in his study on magnitude sound symbolism. Since our results demonstrate a direct match between vowel intrinsic duration and the “size” of musical notes, there is no need to explain the “size” of musical notes via Ohala’s “frequency code” hypothesis. Therefore, a possible answer to the question What is, for example, so small about *mil* and large about *mal*? (Vainio 2021, p. 2) might be: Small about *mil*, is the small intrinsic duration of the vowel /i/, and large about *mal* is the large intrinsic duration of the vowel /a/.

Vowels and a Sound-Symbolic Musical Protolanguage

The non-arbitrary associations between vowel *intrinsic duration* and musical notes are consistent with the results of previous studies (Fenk-Oczlon and Fenk, 2009a,b) reporting non-arbitrary associations between vowel *intrinsic pitch* and musical pitch in meaningless syllables: In songs containing strings of meaningless syllables, vowels are connected to melodic direction in close correspondence to their *intrinsic pitch* or the frequency of the second formant F2. The tight relationships between vowel acoustics and musical intervals indicate that in the case of singing senseless syllables, where there is no pressure of text, vowels and melody seem to merge. This might strengthen the idea that both music and speech evolved from a common prosodic precursor.

In Fenk-Oczlon (2017) I speculated that the earliest human vocal communication may have started with vowels or vowel syllables strung together, which were connected by semivowels or glides such as [w], [h], [j] or the glottal stop [ʔ]. The vowel sequences exhibited pitch and timbre modulations which were used to express different social and pragmatic functions, and were probably propositionally meaningless. The main arguments for this speculation were based on findings from language ontogeny, ethnomusicology, and parallels between vowels and musical patterns. In the 2017 paper I did not consider the huge sound symbolic potential of vowels and their disproportionate role in talker identity discrimination, including characteristics such as age, biological sex, origin, or emotional state. Considering all these properties of vowels, it seems plausible that the sequences of vowel syllables were not *bare phonology* in the sense of Fitch (2010), but instead conveyed sound symbolic information about the environment, about emotional states, or speaker identity. The sequences of vowel syllables probably also contained interjections similar to present-day words such as *ah*, *oh*, *eh*, *huh*. In this context it is interesting to note that Dingemanse et al. (2013) reported that all variants of the interjection word *huh* in their cross-linguistic sample consisted either of a vowel-only syllable, a syllable with a glottal stop [ʔ], or a glottal fricative [h] in the onset.

The vowel sequences were likely very polysemous, because of the small number of vowels (present-day languages have on average 5–6 vowels; Maddieson, 2005) which does not allow much variation in a sequence. Only pitch, duration, intonational contour, rhythmic grouping and situational context could help to discriminate the different (sound symbolic) meanings.

Even in present-day languages, vowel-only sentences can be observed. **Table 1** gives some examples from Japanese (Tsunoda, 1985), Carinthian (my own native knowledge) and vowel-only

expletives from the Mbendjele Pygmies (Lewis, 2009). I am not able to analyze the Japanese examples, but the Carinthian example shows that the word “a”/ a/ is quite polysemous: It can be a question particle, an interjection of astonishment, and also denotes *auch* “also.” The expletives from the Mbendjele Pygmies nicely demonstrate the potential of vowels to convey emotional content. Furthermore, Lewis (2009) reports that vowel-only sentences can also be observed in very intimate communication situations between two persons of the Mbendjele Pygmies, who “tend to omit consonants, leaving only tone and vowels” (Lewis 2009, p. 241).

One might speculate that the earliest stage of human vocal communication, where mere vowel syllables connected by semivowels were strung together, best represents the hypothesized common prosodic precursor of speech and music. The vowel syllables exhibited all core elements of music, pitch, timbre, duration, and intensity. They conveyed prosodic information such as intonation, rhythm, tempo, but also (semantic) sound-symbolic or onomatopoeic information about the environment, inner mental states or speaker identity. In a later stage, consonants such as obstruents emerged and were combined with vowels into consonant-vowel syllables. This was

likely the emergence of articulated speech (Jordania, 2006), and of utterances which could express propositional meaning.

Grauer (2006) speculated that yodeling might be a vestige of the earliest singing style of humanity. The Alpine yodel syllables investigated in this paper may not be too different from the vowel syllables in the hypothesized earliest stage of human vocal communication.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

I thank the reviewers for their insightful comments and helpful suggestions.

REFERENCES

- Bannan, N. (2008). Language out of music: the four dimensions of vocal learning. *Aust. J. Anthropol.* 19, 272–293. doi: 10.1111/j.1835-9310.2008.tb00354.x
- Bentley, M., and Varon, E. J. (1933). An accessory study of “phonetic symbolism.” *Am. J. Psychol.* 45, 76–86. doi: 10.2307/1414187
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proc. Natl. Acad. Sci. U.S.A.* 113, 10818–10823. doi: 10.1073/pnas.1605782113
- Brown, S. (2000). “The Musilanguage model of music evolution,” in: *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press). doi: 10.7551/mitpress/5190.001.0001
- Brown, S. (2017). A joint prosodic origin of language and music. *Front. Psychol.* 8:1894. doi: 10.3389/fpsyg.2017.01894
- Butler, S. (2015). *The Ancient Phonograph*. Boston, MA: Zone Books. doi: 10.2307/j.ctv14gppj13
- Cole, R., Yan, Y., Mak, B., Fenty, M., and Bailey, T. (1996). “The contribution of consonants versus vowels to word recognition in fluent speech,” in: *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing ICASSP’96*. Atlanta, GA.
- Cuskley, C., Dingemans, M., Kirby, S., and van Leeuwen, T. M. (2019). Cross-modal associations and synesthesia: categorical perception and structure in vowel–color mappings in a large online sample. *Behav. Res. Methods* 51, 1651–1675. doi: 10.3758/s13428-019-01203-7
- Cuskley, C., Simner, J., and Kirby, S. (2017). Phonological and orthographic influences in the boba–kiki effect. *Psychol. Res.* 81, 119–130. doi: 10.1007/s00426-015-0709-2
- Cutler, A., and Mehler, J. (1993). The periodicity bias. *J. Phonetics* 21, 103–108. doi: 10.1016/S0095-4470(19)31323-3
- Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: J. Murray. doi: 10.5962/bhl.title.24784
- Dingemans, M., Torreira, F., and Enfield, N. J. (2013). Is ‘Huh?’ a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PLoS ONE* 8:e78273. doi: 10.1371/journal.pone.0078273
- Fenk-Oczlon, G. (2017). What vowels can tell us about the evolution of music. *Front. Psychol.* 8:1581. doi: 10.3389/fpsyg.2017.01581
- Fenk-Oczlon, G., and Fenk, A. (2009a). “Musical pitch in nonsense syllables: correlations with the vowel system and evolutionary perspectives,” in *Proceedings of 7th Triennial Conference of the European Society for the Cognitive Sciences of Music*, eds J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, and P.-S. Eerola (Jyväskylä: European Society for the Cognitive Sciences of Music).
- Fenk-Oczlon, G., and Fenk, A. (2009b). Some parallels between language and music from a cognitive and evolutionary perspective. *Music. Sci.* 13, 201–226. doi: 10.3389/fmns.2016.00274
- Filippi, P. (2020). Emotional voice intonation: A communication code at the origins of speech processing and word-meaning associations? *J. Nonverb. Behav.* 44, 395–417. doi: 10.1007/s10919-020-00337-z
- Filippi, P., and Gingras, B. (2018). “Emotion communication in animal vocalizations, music and language: An evolutionary perspective,” in: *The Talking Species*, eds E. M. Luef and M. M. Marin (Graz: Uni-Press Graz Verlag GmbH).
- Fitch, W. T. (2005). The evolution of language: A comparative review. *Biol. Philosophy* 20, 193–230. doi: 10.1007/s10539-005-5597-1
- Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi: 10.1016/j.cognition.2005.11.009
- Fitch, W. T. (2010). *Evolution of Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511817779
- Fonagy, I. (1981). “Emotions, voice and music,” in: *Research Aspects on Singing*, ed J. Sundberg (Stockholm and Paris: Royal Swedish Academy of Music).
- Fürniss, S. (1991). *Die Jodeltechnik der Aka-Pygmäen in Zentralafrika*. Berlin: Dieter Reimer.
- Grauer, V. A. (2006). Echoes of our forgotten ancestors. *World Music* 48, 5–58.
- Halle, M., Hughes, G. W., and Radley, J. -P. A. (1957). Acoustic properties of stop consonants. *J. Acoust. Soc. Am.* 29, 107. doi: 10.1121/1.1908634
- Haynie, H., Bowern, C., and La Palombara, H. (2014). Sound symbolism in the languages of Australia. *PLoS ONE* 9:e92852. doi: 10.1371/journal.pone.0092852
- House, A. S., and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J. Acoust. Soc. Am.* 25, 105–113. doi: 10.1121/1.1906982
- Jespersen, O. (1933). *Symbolic value of the vowel i. In Linguistica; Selected Papers in English, French, and German*. Copenhagen: Levin & Munksgaard.
- Johansson, N., and Zlatev, J. (2013). Motivations for sound symbolism in spatial deixis: a typological study of 101 languages. *Public J. Semiot.* 5, 3–20. doi: 10.37693/pjos.2013.5.9668

- Johansson, N. E., Anikin, A., Carling, G., and Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguist. Typol.* 24, 253–310. doi: 10.1515/lingty-2020-2034
- Jordania, J. (2006). *Who Asked the First Question? The Origins of Human Choral Singing, Intelligence, Language and Speech. The Origins of Human Choral Singing, Intelligence.* Tbilisi: Logos.
- Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *J. Acoust. Soc. Am.* 122, 2365–2375. doi: 10.1121/1.2773986
- Knoeferle, K., Li, J., Maggioni, E., and Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Sci. Rep.* 7, 1–11. doi: 10.1038/s41598-017-05965-y
- Köhler, W. (1929). *Gestalt Psychology*. New York, NY: Liveright.
- Kolinsky, R., Pascale Lidji, P., Peretz, I., Besson, M., and Morais, J. (2009). Processing interactions between phonology and melody: Vowels sing but consonants speak. *Cognition* 112, 1–20. doi: 10.1016/j.cognition.2009.02.014
- Ladefoged, P. (2001) *Vowels and Consonants: An Introduction to the Sounds of Languages*. Oxford: Blackwell, Blackwell Publications, Malden.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: The MIT Press.
- Levman, B. (1992). The genesis of music and language. *Ethnomusicology* 36, 147–117 doi: 10.2307/851912
- Lewis, J. (2009). “As well as words: Congo Pygmy hunting, mimicry, and play,” in: *The Cradle of Language*, eds R. Botha and C. Knight (Oxford: Oxford University Press).
- Lidji, P., Jolicoeur, P., Régine Kolinsky, R., Moreau, P., Connolly, J. F., and Peretz, I. (2010). Early integration of vowel and pitch processing: A mismatch negativity study. *Clin. Neurophysiol.* 121, 533–541. doi: 10.1016/j.clinph.2009.12.018
- Maddieson, I. (2005). “Vowel quality inventories,” in *The World Atlas of Language Structures*, eds M. Haspelmath, M. S. Dryer, D. Gil, and B. Comrie (Oxford: Oxford University Press).
- Marks, L. E. (1975). On colored-hearing synesthesia: Cross-modal trans-lations of sensory dimensions. *Psychol. Bullet.* 82, 303–331. doi: 10.1037/0033-2909.82.3.303
- Meyer, E. A. (1896). Zur tonbewegung des vokals im gesprochenen und gesungenen einzelwort. *Phonet. Stud.* 10, 1–21.
- Moos, A., Smith, R., Miller, S. R., and Simmons, D. R. (2014). Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder. *i-Perception*, 5, 132–142. doi: 10.1068/i0626
- Nettl, B. (1954). Text-music relationships in Arapaho songs. *Southwestern J. Anthropol.* 10, 192–199. doi: 10.1086/soutjanth.10.2.3628825
- Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41, 1–16. doi: 10.1159/000261706
- Ohala, J. J. (1994). “The frequency code underlies the sound-symbolic use of voice pitch,” in: *Sound Symbolism*, eds H. Leanne, N. Johanna and O. John (Cambridge: Cambridge University Press). doi: 10.1017/CBO9780511751806.022
- Owren, M. J., and Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *J. Acoust. Soc. Am.* 119, 1727–1739. doi: 10.1121/1.2161431
- Parise, C., and Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experi. Brain Res.* 220, 319–333. doi: 10.1007/s00221-012-3140-6
- Patak, A., and Calvert, G. A. (2021). Sooo sweet! Presence of long vowels in brand names lead to expectations of sweetness. *Behav. Sci.* 11:12. doi: 10.3390/bs11020012
- Peña, M., Mehler, J., and Nespore, M. (2011). The role of audiovisual processing in early conceptual development. *Psychol. Sci.* 22, 1419–1421. doi: 10.1177/0956797611421791
- Perlman, M., and Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture* 14, 320–350. doi: 10.1075/gest.14.3.03per
- Peterson, G. E., and Barney, H. L. (1952). Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175–184. doi: 10.1121/1.1906875
- Peterson, G. E., and Lehiste, I. (1960). Duration of syllable nuclei in English. *J. Acoustical Soc. Am.* 32, 693–703. doi: 10.1121/1.1908183
- Pommer, J. (1906). *Zwanzig echte alte Jodler*. Wien: Adolf Robitschek.
- Rabaglia, C. D., Maglio, S. J., Krehm, M., Seok, J. H., and Trope, Y. (2016). The sound of distance. *Cognition* 152, 141–149. doi: 10.1016/j.cognition.2016.04.001
- Ramachandran, V. S., and Hubbard, E. M. (2001). Synaesthesia – a window into perception, thought and language. *J. Consciousness Stud.* 8, 3–34.
- Reybrouck, M., and Podlipniak, P. (2019). Preconceptual spectral and temporal cues as a source of meaning in speech and music. *Brain Sci.* 9:53. doi: 10.3390/brainsci9030053
- Rousseau, J.-J. (1781). *Essay on the Origin of Languages. English Translation by J. H. Moran and A. Gode (1986)*. Chicago, IL: University of Chicago Press.
- Rummer, R., Schweppe, J., Schlegelmilch, R., and Grice, M. (2014). Mood is linked to vowel type: The role of articulatory movements. *Emotion* 14, 246–250. doi: 10.1037/a0035752
- Sapir, E. (1929). A study in phonetic symbolism. *J. Experi. Psychol.* 12:225. doi: 10.1037/h0070931
- Scherer, K. R. (1995). Expression of emotion in voice and music. *J. Voice* 9, 235–248. doi: 10.1016/S0892-1997(05)80231-0
- Shinohara, K., and Kawahara, S. (2016). “A cross-linguistic study of sound symbolism: the images of size,” in *Proceedings of the Thirty-Sixth Annual Meeting of the Berkeley Linguistics Society*. Berkeley. doi: 10.3765/bls.v36i1.3926
- Simner, J., Cuskley, C., and Kirby, S. (2010). What sound does that taste? Cross-modal mappings across gustation and audition. *Perception* 39, 553–569. doi: 10.1068/p6591
- Solé, M. J., Ohala, J. J. (2010). “What is and what is not under the control of the speaker. Intrinsic vowel duration,” in: *Papers in Laboratory Phonology 10*, eds C. Fougerson, B. Kühnert, M. D’Imperio, and N. Vallée (Berlin: de Gruyter).
- Thompson, W. F., Marin, M. M., and Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19027–19032. doi: 10.1073/pnas.1210344109
- Toivonen, I., Blumenfeld, L., Gormley, A., Hoiting, L., Lo-gan, J., Ramlakhan, N., and Stone, A. (2015) “Vowel height and duration,” *Proceedings of the 32nd West Coast Conference on Formal Linguistics*.
- Traunmüller, H. (1986). “Some aspects of the sound of speech sounds,” in *The Psychophysics of Speech Perception*, ed M. E. H. Schouten (Dordrecht: Martinus Nijhoff), 293–305. doi: 10.1007/978-94-009-3629-4_24
- Tsunoda, T. (1985). *The Japanese Brain*. Tokyo: Taishukan.
- Ullian, R. (1978). “Size-sound symbolism,” in *Universals of Human Language: Phonology*, eds J. Greenberg (Stanford, CA: Stanford University Press).
- Vainio, L. (2021). Magnitude sound symbolism influences vowel production. *J. Memory Lang.* 118:104213. doi: 10.1016/j.jml.2020.104213
- Wey, Y. (2019). *Transkription wortloser Gesänge*. Innsbruck: Innsbruck University Press. doi: 10.15203/3187-81-8
- Whalen, D. H., Levitt, A. G., Hsiao, P.-L., and Smorodinsky, I. (1995). Intrinsic F0 of vowels in the babbling of 6-, 9- and 12-month-old French-and English-learning infants. *J. Acoustical Soc. Am.* 97, 2533–39. doi: 10.1121/1.411973
- Winter, B., and Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa J. Gen. Linguist.* 6, 1–13. doi: 10.5334/gjgl.1646
- Zhang, C., Shao, J., Huang, X. (2017). Deficits of congenital amusia beyond pitch: Evidence from impaired categorical perception of vowels in Cantonese-speaking congenital amusics. *PLoS ONE* 12:e0183151. doi: 10.1371/journal.pone.0183151

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Fenk-Oczlon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.