




Cognitive Science 44 (2020) e12906

© The Authors. *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS). All rights reserved.

ISSN: 1551-6709 online

DOI: 10.1111/cogs.12906

Sound Predicts Meaning: Cross-Modal Associations Between Formant Frequency and Emotional Tone in Stanzas

Jan Auracher,^a  Winfried Menninghaus,^a Mathias Scharinger^b

^a*Department of Language and Literature, Max Planck Institute for Empirical Aesthetics*

^b*Institute for German Linguistics, Philipps University of Marburg*

Received 14 December 2019; received in revised form 12 August 2020; accepted 18 August 2020

Abstract

Research on the relation between sound and meaning in language has reported substantial evidence for implicit associations between articulatory–acoustic characteristics of phonemes and emotions. In the present study, we specifically tested the relation between the acoustic properties of a text and its emotional tone as perceived by readers. To this end, we asked participants to assess the emotional tone of single stanzas extracted from a large variety of poems. The selected stanzas had either an extremely high, a neutral, or an extremely low average formant dispersion. To assess the average formant dispersion per stanza, all words were phonetically transcribed and the distance between the first and second formant per vowel was calculated. Building on a long tradition of research on associations between sound frequency on the one hand and non-acoustic concepts such as size, strength, or happiness on the other hand, we hypothesized that stanzas with an extremely high average formant dispersion would be rated lower on items referring to *Potency* (dominance) and higher on items referring to *Activity* (arousal) and *Evaluation* (emotional valence). The results confirmed our hypotheses for the dimensions of Potency and Evaluation, but not for the dimension of Activity. We conclude that, at least in poetic language, extreme values of acoustic features of vowels are a significant predictor for the emotional tone of a text.

Keywords: Sound iconicity; Formant dispersion; Phonosemantics; Sound symbolism; Emotional tone; Sound-meaning congruence

Correspondence should be sent to Jan Auracher, Kyoto-Women's University, 35 Kitahiyoshicho, Higashi-yama-ku, 605-8501 Kyoto, Japan. E-mail: auracher@kyoto-wu.ac.jp

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Interest in the relation between sound and meaning in human language dates back more than 2,000 years to the Greek philosopher Plato (ca. 360 BC) and has since been discussed by many renowned figures, such as Lomonosov (1748), Locke (1690), Leibniz (1765), Humboldt (1836), Jespersen (1922), and Sapir (1929) (for reviews of the history of research on sound iconicity, see, e.g., Allott, 1995; Jakobson & Waugh, 2002; Kjetsaa, 1974; Magnus, 2001; Nuckolls, 1999). The empirical research of this topic began in the early 20th century and has recently gained momentum (for reviews, see Akita, 2009; Hinton, Nichols, & Ohala, 1994; Lockwood & Dingemans, 2015; Monaghan, Shillcock, Christiansen, & Kirby, 2014; Perniss, Thompson, & Vigliocco, 2010; Schmidtko, Conrad, & Jacobs, 2014; Sidhu & Pexman, 2018). However, there is still no consistently used terminology for different kinds of phonosemantic relations. In what follows, we use the term *sound iconicity* to designate cross-modal associations between acoustic properties of phonemes and non-acoustic attributes, such as large, fast, or happy (Jakobson & Waugh, 2002, p. 196; Reay, 1994; Wescott, 1971). A similar definition for sound iconicity has recently been introduced by Dingemans, Blasi, Lupyan, Christiansen, and Monaghan (2015). However, in contrast to the approach followed in this study, Dingemans et al. relate iconicity to any kind of perceptuomotor analogy, thus, also including onomatopoeia (i.e., sound–meaning relations based on imitation). Dingemans et al. further distinguish between different subcategories of iconicity. According to this terminology, our definition of sound iconicity is closest to what they refer to as *relative* or *diagrammatic iconicity* (Dingemans et al., 2015, p. 606).

Whereas most studies to date on sound iconicity have focused on providing proofs for the very existence of phonosemantics, research has recently turned to investigating its particular domains, dimensions, and functions. For example, sound iconicity has been found to support language acquisition (Imai & Kita, 2014; Imai, Kita, Nagumo, & Okada, 2008; Kantartzis, Imai, & Kita, 2011; Lockwood, Dingemans, & Hagoort, 2016; Nygaard, Cook, & Namy, 2009; Perry, Perlman, & Lupyan, 2015; Perry, Perlman, Winter, Massaro, & Lupyan, 2018; Yoshida, 2012), to influence the intonation in verbal interactions (Herold, Nygaard, & Namy, 2012; Nygaard, Herold, & Namy, 2009; Perlman, Clark, & Johansson Falck, 2015), and to be used in brand names to foster associations between a product and its desirable attributes (Coulter, 2009; Coulter & Coulter, 2010; Duduciuc, 2015; Klink, 2000; Lowrey & Shrum, 2007; Pogacar, Plant, Rosulek, & Kouril, 2015; Shrum, Lowrey, Luna, Lerman, & Liu, 2012; Yorkston & Menon, 2004). Moreover, evidence suggests that there is a statistically measurable relation between the meaning of words and their phonetic properties (English: Heise, 1966; Nastase, Sokolova, & Shirabad, 2007; Whissell, 1999; German: Aryani, Conrad, Schmidtko, & Jacobs, 2018; Fónagy, 1963) and that this applies across various languages from different language families (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016; Dautriche, Mahowald, Gibson, & Piantadosi, 2017; Ultan, 1978).

One branch of studies within the realm of sound iconicity has investigated the relationship between the content of texts and the frequency of occurrence of particular phonemes

in these texts. Compared to other areas of research on sound iconicity, the results from these studies provide a fairly ambiguous picture. Whereas a majority of the studies have reported significant relations between the emotional tone of texts and the relative occurrence of certain phonemes (Aryani, Conrad, & Jacobs, 2013; Auracher, Albers, Zhai, Gargeeva, & Stavniychuk, 2010; Auracher, Scharinger, & Menninghaus, 2019; Fónagy, 1961; Whissell, 1999, 2003, 2004a, 2004b, 2011, 2016, 2017, 2018), other studies have found no such relations (Kraxenberger & Menninghaus, 2016; Miall, 2001). What is more, the results of the studies that did find significant sound-iconic relations vary greatly and even seem to partially contradict each other (Tsur & Gafni, 2020).

We propose that these inconsistencies can partly be explained by differences in the applied research designs (Auracher, 2020). Starting with Humboldt, it has repeatedly been suggested that there are various kinds of phonosemantic relations in language that differ regarding the underlying cognitive processes (Humboldt, 1836; see also Magnus, 2001). A detailed discussion of the different kinds of phonosemantic relations would clearly go beyond the scope of this article (for reviews, see Hinton et al., 1994; Perniss et al., 2010; Reay, 1994; Schmidtke et al., 2014; Sidhu & Pexman, 2018). However, we would like to point out that our decision for the research design applied in this study was motivated by the specific nature of the phonosemantic relation at stake and might, therefore, not be comparable to results of other studies with a different research interest.

First, while most studies so far have focused on the relative occurrence of individual phonemes as a function of a text's content, this study will focus on the relation between the content of a text and its acoustic characteristics. There is good evidence for affective associations with either individual phonemes or the articulatory–acoustic properties of phonemes. For example, Monaghan and Fletcher (2019), asking participants to assess the meaning of non-words, found that statistical models based on individual phonemes fitted participants' responses significantly better than models based on phonetic features. Aryani and colleagues, applying neuroimaging techniques (fMRI), reported that words that sounded arousing led to increased activity in areas of the brain that also respond to non-verbal emotional expressions and emotional prosody (Aryani, Hsu, & Jacobs, 2018). Additionally, Aryani, Conrad, et al. (2018) reported that acoustic characteristics of words can account for a significant portion of their affective meaning and that these acoustic characteristics could also predict the assessment of the emotional valence and level of arousal of pseudowords. Comparing the effects of articulatory and acoustic features of phonemes on sound-iconic associations, Ohtake and Haryu (2013) observed that the effect of acoustic characteristics was independent from that of articulatory movements, whereas articulatory characteristics alone had no effect on sound-iconic associations. What is more, it has been claimed that some forms of sound iconicity are simply a specific variant of cross-modal associations, that is, of associations between basic features of perception that address different sensory modalities (e.g., Marks, 1978; Ohala, 1994; Parise, Knorre, & Ernst, 2014; Parise & Spence, 2013; Spence, 2011). The aim of the present study, therefore, is to test the role of such cross-modal associations for the perception of written language. Consequently, we will study the relation between the *acoustic features* of phonemes and a text's content rather than counting the frequency of distinct phonemes. Note

that by *acoustic characteristics of phonemes*, we refer to the average values of acoustic properties of phonemes that have been measured for a certain number of people. That is, individual characteristics of the voice of a speaker, effects of emotional prosody, or modulations of a phoneme's pronunciation due to the specific phonetic constellation in a word will not be taken into account.

Second, a majority of the studies on sound iconicity in written language were conducted using what is, on our understanding, an insufficient number of text samples. We believe that it is necessary to draw on a large number of text samples in order to isolate the potential effects of sound iconicity from the influence of other linguistic features.

Third, in this study, rather than treating the relative occurrence of acoustic features in a text as a dependent variable, we test the power of acoustic features to predict readers' emotional assessment of texts (Auracher et al., 2010).

Finally, earlier studies often applied ad hoc categories to assess sound-iconic associations of phonemes or phonetic features. In contrast, for this study, we employ a well-established dimensional model of emotions comprising the three factors *Evaluation*, *Potency*, and *Activity* (the EPA model) as introduced by Osgood, Suci, and Tannenbaum (1957). According to this model, an emotion can be defined by its position in a three-dimensional space that characterizes states as positive or negative (Evaluation), dominant or submissive (Potency), and active or passive (Activity). Later versions of the EPA model have often referred to the three dimensions as *Valence* or *Pleasure* (for Evaluation), *Arousal* or *Intensity* (for Activity), and *Dominance* or *Controllability* (for Potency) (e.g., Mehrabian & Russell, 1974; Trnka, Lačev, Balcar, Kuška, & Tavel, 2016). While we consider these terms as largely interchangeable, it is important to point out some nuances that set them apart. Bakker and colleagues have noted that the terms used in these models sometimes differ regarding their focus on affect, cognition, or behavior (Bakker, van der Voordt, Vink, & de Boon, 2014). For example, while Osgood et al. (1957) operationalized the Activity dimension in their EPA model by using antonym pairs such as *fast-slow* or *active-passive*, Mehrabian and Russell (1974) replaced the term Activity with Arousal in their PAD-model, and, in so doing, shifted the focus from physical or cognitive activity to an inner state. Although the different stages related to Activity or Arousal are potentially interrelated, it is also true that the flexibility of the EPA model makes it conceptually fuzzy and its dimensions hard to define. Consequently, results of studies that used the EPA dimensions are often difficult to compare, as the operationalization of the dimensions can show substantial variations. For our study, we decided to stick to the original terms of the EPA model. However, we will also use different terms when reporting results of other studies depending on the focus of their respective research interests.

1.1. *Research objective and hypothesis*

Research on sound-iconic properties of vowels has mostly focused on the effect of their articulatory place (i.e., front vs. back vowels) and/or articulatory height (i.e., high or closed vowels vs. low or open vowels) on cross-modal associations (for a broader

overview, see Material S1). The majority of these studies have addressed the phonosemantic relation between articulatory features of vowels and the notion of size, strength, or dominance more generally (for reviews, see Auracher, 2017; Davis, Morrow, & Lupyan, 2019; Hoshi, Kwon, Akita, & Auracher, 2019; Nuckolls, 1999; Ohala, 1994; Sidhu & Pexman, 2018; Tsur, 2006). In sum, there is strong evidence suggesting that low-back vowels are preferably associated with high level of Potency or Dominance whereas high-front vowels are mostly related to low Potency or Submissiveness. As articulatory features of vowels are directly linked to their acoustic characteristics, the results of these studies can also be translated into sound-iconic associations. That is, the place of articulation has a strong influence on the frequency of the second formant, whereas the height of articulation affects the frequency of the first formant (Ladefoged & Disner, 2012). As a consequence, the distinction between high front vowels, such as /i:/ in *free*, and low back vowels, such as /a:/ in *spa*, is mirrored in the relative formant dispersion, that is, the vowel's characteristic distance between the first and second formant (Fig. 1).

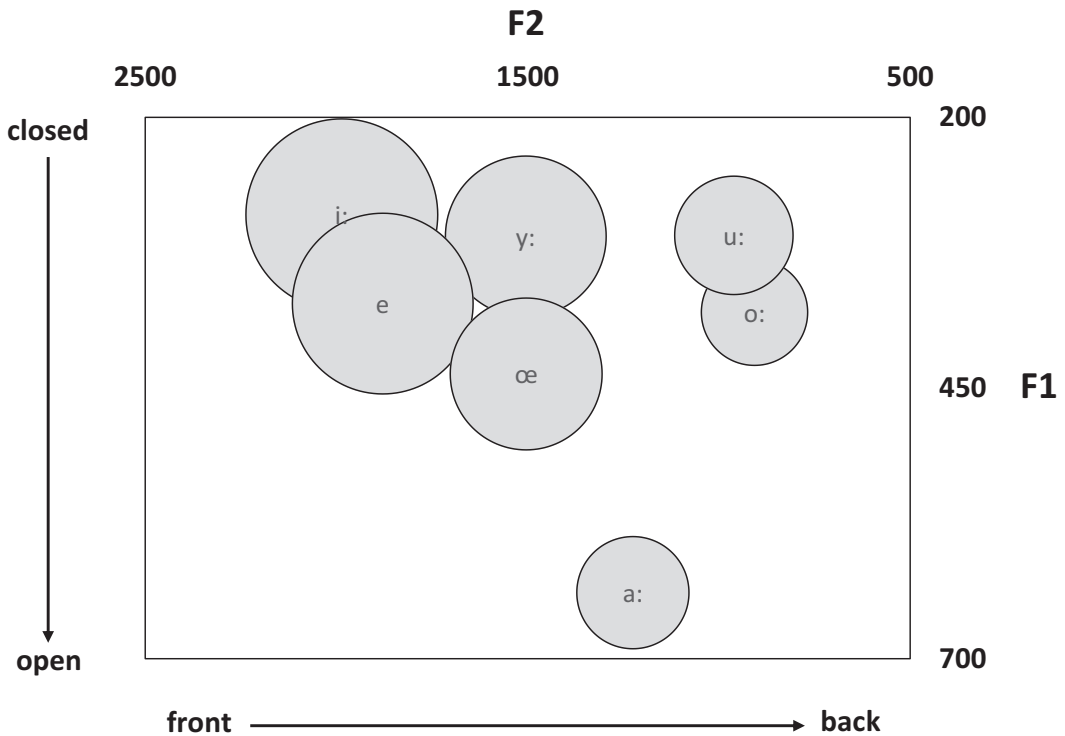


Fig. 1. Values of the first two formants and formant dispersion for a selection of German vowels. The size of the bubbles indicates the values for the formant dispersion. The first formant (F1) increases from closed vowels to open vowels (vertical axis), and the second formant (F2) increases from back vowels to front vowels (horizontal axis). The formant dispersion decreases from closed front vowels (/i:/) to back vowels (/u:/, /o:/) and open vowels (/a:/) (top left to bottom right).

Compared to the well-studied relation between vowels' articulatory features and Potency, considerably fewer studies have tested associations between vowels and concepts related to Evaluation or Activity. For Evaluation, the studies by Klink (2000), Markel and Hamp (1960), Miron (1961), and Tarte (1982) provided evidence that front vowels (compared to back vowels) received higher ratings for *pleasant*, *beautiful*, *harmonious*, and *good*. Whereas these three studies employed questionnaires to directly test sound-iconic associations with vowels' articulatory-acoustic features, recent studies applied more indirect approaches. For example, some studies considered the relative occurrence of phonemes or phonetic-acoustic features in annotated wordlists (i.e., wordlists with ratings for Valence and Arousal per word) as an indicator for the emotional connotations of these phonemes (e.g., Adelman, Estes, & Cossu, 2018; Aryani, Conrad, et al., 2018; Louwerse & Qu, 2017). Similar strategies to tackle associations between phonemes (or phonetic features) and emotional tone have previously been applied by Ertel (1965) and Heise (1966). However, neither of these studies reported a specific relation between formant dispersion (the way it is operationalized in our study) and the dimensions of Evaluation or Activity.

Taking a different approach, Meyers and colleagues found that dynamic upward or downward shifts of the formant frequencies (particularly of the second formant) predict emotional associations with pseudo-words in a forced-choice experiment (Myers-Schulz, Pujara, Wolf, & Koenigs, 2013). These findings can also be interpreted as supporting a positive correlation between Evaluation and formant dispersion, given that downward shifts of the second formant, which have been found to be associated with negative emotional valence, diminish the distance between the first and second formant and, therefore, the formant dispersion. Conversely, upward shifts, which increase the formant dispersion, were more readily matched with positive emotional valence. However, the exact relation between emotional associations of dynamic shifts of formant frequencies on the one hand and those of static measurements of formant dispersion on the other hand remains a question for further studies.

A specific relation between the vowels /i:/ (high-front vowel) and /o:/ (back-vowel) and emotional valence has been demonstrated by Rummer, Schweppe, Schlegelmilch, and Grice (2014). The authors reported that participants who repeatedly articulated the phoneme /i:/ while watching a cartoon rated it as funnier than participants who repeatedly articulated the phoneme /o:/. The authors explain this in terms of the parallels between emotional facial gestures that express positive and negative mood with the articulatory movements of the vowels /i:/ and /o:/, respectively.

While only few studies have tested sound-iconic associations between vowels and Evaluation, research on the relation between vowels' articulatory-acoustic characteristics and lightness or brightness has some potential to compensate for this lack of direct empirical evidence. Vowels with a high formant dispersion have repeatedly been found to be closely related to brightness, while vowels with a low formant dispersion have been found to be more readily associated with darkness (Becker & Fisher, 1988; Bentley & Varon, 1933; Fischer-Jørgensen, 1968; Greenberg & Jenkins, 1966; Markel & Hamp, 1960; Newman, 1933). Similarly, studies on cross-modal associations between vowels

and colors have reported that bright colors are preferably matched with high front vowels, while dark colors are usually related to low back vowels (Cuskley, Dingemans, Kirby, & van Leeuwen, 2019; Moos, Smith, Miller, & Simmons, 2014; Park & Osera, 2008; Wrembel & Rataj, 2008). Given that in most cultures the opposition between bright and dark is also metaphorically used to express emotional valence (e.g., *dark mood* [Eng.], *strahlendes Lächeln* [German], *akarui hito* [Japanese], and so forth; see also Adams & Osgood, 1973; Crawford, 2009; Meier, Robinson, Crawford, & Ahlvers, 2007), we assume that the relation between vowels and brightness accounts at least for a significant fraction of the associations between vowels and (positive) Evaluation.

Regarding the relation between vowels and Activity, front vowels (as compared to back vowels) have been found to be preferably assessed as “fast” (Folkins & Lenrow, 1966; Klink, 2000; Klink & Wu, 2014; Lowrey & Shrum, 2007; Markel & Hamp, 1960; Tarte, 1982), “active” (Folkins & Lenrow, 1966; Greenberg & Jenkins, 1966; Markel & Hamp, 1960; Tarte, 1982), and “tense” (Markel & Hamp, 1960), suggesting that a high formant dispersion is associated with a high degree of activity and tension. However, these studies mostly focused on associations of articulatory–acoustic characteristics of vowels with a notion of activity rather than with an inner state of arousal. In contrast, other studies have shown that a speaker’s level of arousal has an effect on formant frequency, with higher arousal leading to a higher first formant (Goudbeek, Goldman, & Scherer, 2009; Laukka, Juslin, & Bresin, 2005; Vlasenko, Prylipko, Philippou-Hübner, & Wendemuth, 2011). However, as the frequency of the first formant alone is not sufficient to draw conclusions on the formant dispersion, that is, the distance between the first two formants, it is not clear how these results relate to our study. A similar problem arises when comparing our study with the one by Aryani, Conrad, et al. (2018) who tested the relation between Arousal and formant bandwidths as well as between Arousal and standard deviations of spectral center of gravity. These measurements not only contrast with our definition of formant dispersion (distance between mean F2 and F1), thereby precluding a direct comparison, but also introduce acoustic parameters that are prone to be dependent on acoustic context and speaker properties—variables that have been excluded in our study. Thus, to the best of our knowledge, there are currently no studies that tested the relation between formant dispersion, defined as the distance between the first two formants, and arousal, defined as an affective state.

Taken together, studies that monitored the associations of articulatory and acoustic properties of vowels with emotions suggest a relation between high front vowels, that is, vowels with a high value for formant dispersion, and an emotional tone that is characterized by a positive valence, a feeling of weakness or submissiveness, and an active or aroused state. In contrast, low back vowels, that is, vowels with a low formant dispersion, should be associated with a negative emotional valence, dominance, and calmness. However, as outlined above, findings regarding the relation between arousal and the frequency of the first formant indicate that the relation between formant dispersion and Activity is not clear yet and calls for further research. Hypothesizing a close relation between the acoustic properties of a text’s phonemes and its content, we conducted a study in which

we asked participants to assess the content of short text samples categorized according to their average formant dispersion on six bipolar items comprising the EPA model.

2. Materials and methods

2.1. Participants

The text samples were assessed by 43 participants (29 women, 10 men, 4 undisclosed). Participants were recruited from a pool of volunteers who had signed up in the participant database of the Max Planck Institute for Empirical Aesthetics (Frankfurt am Main, Germany). The average age of participants was 25.7 years (min: 20, max: 40, STD: 5.14). A majority (35) of the participants were undergraduate (27) or graduate students (8). The remaining participants were mostly employees (5) or provided no information regarding their current employment situation. All participants were native German speakers and reported normal or corrected-to-normal sight.

2.2. Ethics statement

This study was conducted in full accordance with the World Medical Association's Declaration of Helsinki and the Ethical Guidelines of the German Association of Psychologists (DGPs). All participants gave their written informed consent and received EUR 10 as compensation for their participation, independently of whether or not they finished the study. All of the study procedures were approved by the Ethics Council of the Max Planck Society.

2.3. Materials

Assuming that phonosemantic relations are particularly common in poetic language, we used stanzas from German poems as stimuli. Text samples were selected from an online archive of classical German poems (Freiburger Anthologie, <http://freiburger-anthologie.ub.uni-freiburg.de/>). The archive comprises more than 1,500 German poems written between 1,720 and 1,900. After removing duplicates and poems that were written in languages other than standard German (e.g., in Latin or in a German dialect), our collection consisted of 1,399 poems. From these poems, we selected all stanzas that had a minimum of four and a maximum of six lines. As a result, we had a collection of 8,031 stanzas. All stanzas were phonetically transcribed using the Python library *epitran* (Mortensen, Dalmia, & Littell, 2018). To calculate the average formant dispersion per stanza, we assessed the distance between the first (F1) and second (F2) formant ($dF = |F1 - F2|$) and converted the result from Hertz to Mel (O'Shaughnessy, 1999). The values for the first two formants of each vowel were taken from Pätzold and Simpson (1997; see Material S2). For diphthongs, we followed Greenberg and Jenkins's (1966) analysis and used the values assessed for the latter part of the diphthongs.

For the experiment, we divided the stanzas into three categories. To this end, we assessed an upper and a lower threshold defined by the average formant dispersion of all stanzas plus/minus the standard deviation (mean: 727 Mel, STD: 40 Mel, upper threshold: 767 Mel, lower threshold: 687 Mel). We then randomly selected 90 stanzas, 30 with an extremely high formant dispersion ($dF >$ upper threshold), 30 with an extremely low formant dispersion ($dF <$ lower threshold), and 30 with a formant dispersion between the upper and lower thresholds. In what follows, we refer to these three groups of stanzas as *bright*, *dark*, and *neutral*, respectively (Fig. 2). An overview of the selected stanzas used in the experiment can be found in Material S3.

2.4. Design and procedure

Every stanza was assessed by 14 or 15 participants (mean: 14.3) on six bipolar items, with two items for each of the three EPA dimensions (Evaluation, Potency, Activity; see Table 1). To test the inter-rater agreement, an intraclass correlation coefficient (ICC) per stanza was assessed (Koo & Li, 2016; Shrout & Fleiss, 1979). On average, the inter-rater agreement was *good* with a majority of three-quarters of the coefficients above .73 and

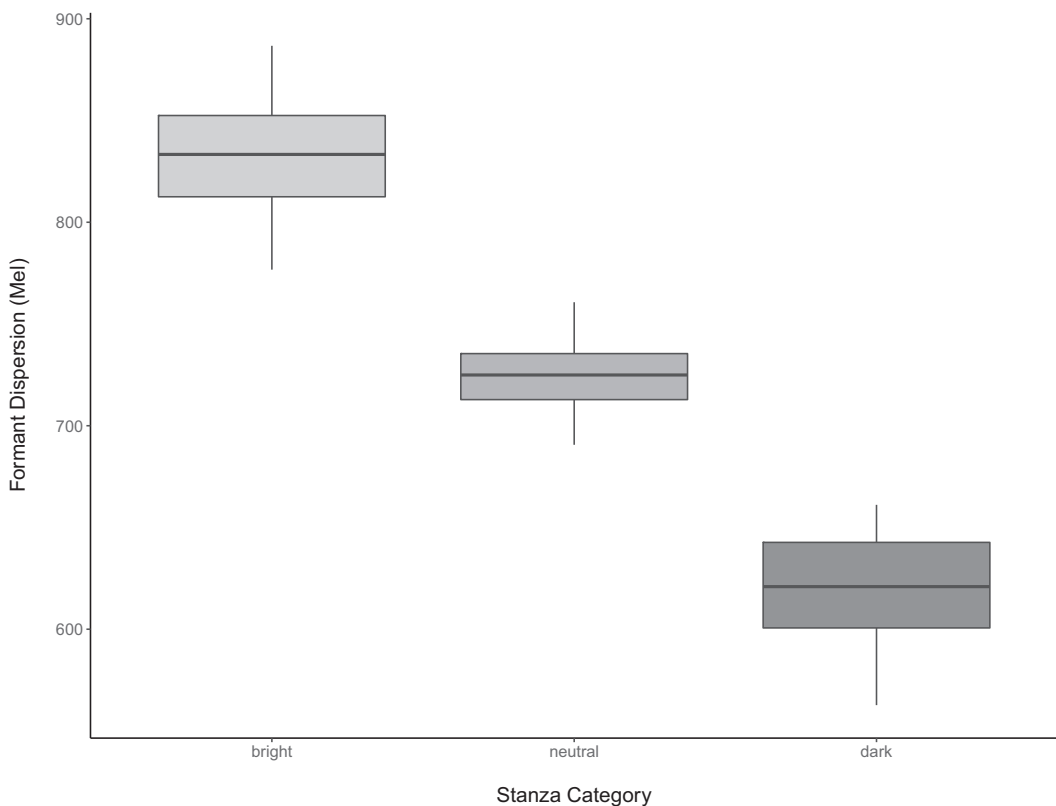


Fig. 2. Median and range of average formant dispersion per category of stanzas.

Table 1

Bipolar items of the three dimensions of the EPA model. Items on the right side represent positive values (+5) and items on the left side represent negative values (−5). The adjectives in italics were hypothesized to be related to a relatively wide formant dispersion (i.e., + Evaluation, − Potency, + Activity). The German designations of the items used in the study are indicated in square brackets

Evaluation		Potency		Activity	
Dark [dunkel]	<i>Bright</i> [hell]	<i>Delicate</i> [zierlich]	Strong [kräftig]	Quiet [ruhig]	<i>Lively</i> [lebhaft]
Sad [traurig]	<i>Happy</i> [fröhlich]	<i>peaceful</i> [friedlich]	Aggressive [aggressiv]	Gentle [sanft]	<i>Vigorous</i> [energisch]

only seven of the coefficients below .5 (N: 90, Md: 0.82, IQR: 0.19, Min: −0.04, Max: 0.98).

As recommended by Osgood et al. (1957), we adapted the items used in this study to fit the specific nature of our stimuli. Specifically, we selected items that clearly indicated that participants were supposed to assess the content of the stanzas and not their effect on the readers. However, as outlined above, the selection of a certain set of items inevitably has the side effect that the overall scope of the underlying dimensions is somewhat narrowed. For example, while most studies use items such as *pleasant* versus *unpleasant* to assess differences regarding the Evaluation dimension, these adjectives imply that the item asks for the effect of a stimulus on the participant. In contrast, we were interested in the participant's evaluation of the emotional tone *expressed in a text*. These two perspectives do not necessarily aim at the same effect. It is possible that a text expresses a negative emotion yet is still perceived as pleasant by the reader. In our study, we, therefore, used the adjective-pair *happy* versus *sad*. Negative or positive emotional valence, however, is not restricted to the opposition between happiness or sadness. For example, there are other negative emotions, such as *contempt*, *anger*, or *fear*, that are not covered by *sadness*. Therefore, due to the specific operationalization of the Evaluation dimension used in this study, participants might have chosen neutral values to assess poems that—though expressing positive or negative emotions—are related to neither *happiness* nor *sadness*.

The poems and rating scales were presented on computer screens using the Tk-inter package of the Python programming environment (<https://docs.python.org/2/library/tkinter.html>). Each item was rated using the Tk-inter package's slider function. Participants were instructed to assess the stanzas by placing the slider closer to one end of the scale or the other (for the precise wording of the instructions, see Material S4). The order of the items and their orientation (i.e., *dark* to the left and *bright* to the right, or vice versa) were randomized per participant.

At the beginning of the experiment, participants were informed about the task and purpose of the study. However, the role of the acoustic properties of the stanzas for our research question was not revealed. Subsequently, participants were asked to assess the stanzas by judging which of the two adjectives for each item appeared to be a closer match to the content. They were also told that, in cases where they believed that neither

of the two adjectives adequately described the stanza, they could set the slider to the middle. Every participant assessed a set of 30 pseudo-randomly selected stanzas (10 per category). The order in which the stanzas were presented was randomized. Before the experiment began, participants practiced the procedure by assessing the first stanza of the poem *Es sitzt ein Vogel auf dem Leim* by the German poet Wilhelm Busch. If they had no further questions, participants started the experiment and went through it at their own pace. At the end of the experiment, participants were asked to provide personal information regarding their gender, age, profession, and educational background. Finally, they were asked for comments. Altogether, the experiment took between 30 and 45 min. All data can be found in Material S5.

2.5. Analysis of the data

To test the effect of the acoustic features of stanzas on readers' assessment, we used a linear mixed regression model that included *stanza categories* (i.e., bright, dark, neutral) as fixed factors and *participant* as a random factor. The random factor was included to account for the fact that each participant assessed only a selection of the stanzas. The data were analyzed using the R software package (R Core Team, 2019). The pairs of items were combined into three dimensions, representing the three-dimensional EPA space. The correlation between the combined items were $r = .66$ for Evaluation, $r = .54$ for Potency, and $r = .64$ for Activity. Linear mixed models conducted for each dimension using the library lme4 (Bates, Mächler, Bolker, & Walker, 2015). To display the results of the analysis, we applied the `tab_model()` function of the sjPlot library (Lüdtke, 2017). For analysis of the data, all values were z -standardized (i.e., $M = 0$ and $SD = 1$).

3. Results

In a first step, we compared the performance of different models with increasing complexity. The results showed that inclusion of random slopes for each participant did not result in a significant improvement (Table 2). To assess the coefficients, we therefore applied a model that included only random intercepts per participant (Model #2). As our hypothesis made distinct predictions for each dimension of the EPA space, we tested three separate models for Evaluation, Potency, and Activity (Table 3).

The results confirmed our hypothesis regarding the Evaluation and the Potency dimensions, but not regarding the Activity dimension: Ratings of bright stanzas are higher for items referring to Evaluation and lower for items referring to Potency (Fig. 3; the exact values for each item are given in Appendix Table A1). Notably, ratings for the neutral stanzas were always close to zero, whereas those for the stanzas with extreme values (i.e., bright and dark) had opposite algebraic signs (+/-). In contrast to our hypothesis, however, ratings for dark stanzas were also higher for Activity. Moreover, while the ratings for the three categories follow the expected order (*bright* < *neutral* < *dark*) for the Potency dimension, the neutral stanza ratings were highest for the Evaluation dimension.

Table 2

Effect of formant dispersion (fixed factor) on readers' assessment of stanzas. Goodness of fit is indicated by the log-likelihood ($-2LL$) criterion. Successive models were compared using the chi-square test

No.	Model	Dimension	$-2LL$	df	χ^2
#1	Dim ~ 1 + 1 Part	Evaluation	3658.6		
		Potency	3655.3		
		Activity	3654.2		
#2	Dim ~ Cat + 1 Part	Evaluation	3641.0	2	17.648***
		Potency	3584.8	2	70.479***
		Activity	3587.6	2	66.556***
#3	Dim ~ Cat + Cat Part	Evaluation	3639.6	5	1.348 ^{ns}
		Potency	3579.3	5	5.480 ^{ns}
		Activity	3579.0	5	8.606 ^{ns}

Note. Stars indicate level of significance with ^{ns} $p > .1$ and *** $p < .001$.

Abbreviations: Cat, category; df , degree of freedom; Dim, dimension; Part, participants.

Looking at the differences between the three categories for each dimension, it is striking that there is always one category of stanzas that stands out compared to the other two categories. For the Evaluation dimension, dark stanzas are clearly different from both bright and neutral stanzas, while bright and neutral stanzas are not significantly different from each other. Conversely, for the Activity dimension, we found a significant difference for bright versus dark and neutral stanzas, but not for dark versus neutral stanzas. Significant differences in both directions (i.e., between neutral and bright stanzas as well as between neutral and dark stanzas) were found only for the Potency dimension. However, here again, the actual effect is clearly stronger for the bright stanzas than for the dark stanzas. This observation is also confirmed by the mixed regression model: For the dimensions of Evaluation and Activity, only one category differs significantly from neutral stanzas.

The ratings for the individual items also followed the expected order. Neutral stanzas were rated with intermediate values compared to the two extreme categories, with the ratings for *happy-sad* [fröhlich-traurig] being the single exception to this rule (see Fig. 4; for exact values, see Appendix Table A2, and for a post hoc comparison of means between the categories, see Material S6). That is, for most items (except *happy-sad*), the ratings of bright and dark stanzas point toward opposite directions, indicating that, on average, stanzas of the two extreme categories were associated with opposite attributes (e.g., bright stanzas with *weak* and dark stanzas with *strong*). As outlined before, the observation that ratings for *happy-sad* are the only exception from this rule might be due to the fact that the opposition between happiness and sadness does not cover the full range of positive and negative emotions.

Our results therefore suggest a close relationship between the average ratings per stanza category (i.e., bright, neutral, and dark stanzas) and the average formant dispersion for these categories. This observation raises the question of whether there is a continuous relationship between the average formant dispersion and the perceived emotional tone of

Table 3
Coefficients (Estimates) for *bright* and *dark* stanzas for each EPA dimension (Model #2: Dim ~ Cat + 1|Part)

Predictors	Evaluation			Potency			Activity		
	Estimates	CI	<i>t</i> values	Estimates	CI	<i>t</i> values	Estimates	CI	<i>t</i> values
Neutral-bright	-0.06	-0.19 - 0.08	-0.85 ^{ns}	-0.38	-0.50 - -0.25	-5.72 ^{***}	-0.41	-0.53 - -0.28	-6.18 ^{***}
Neutral-dark	-0.27	-0.40 - -0.14	-4.00 ^{***}	0.17	0.04 - 0.30	2.61 ^{**}	0.11	-0.02 - 0.24	1.67 ⁺
Bright-dark	-0.21	-0.34 - -0.08	-3.15 ^{**}	0.55	0.42 - 0.68	8.23 ^{***}	0.52	0.39 - 0.64	7.85 ^{***}
Random effects									
σ^2	0.98			0.93			0.93		
ICC	0.01			0.02			0.02		
N	43			43			43		
Observations:	1290			1290			1290		
Marginal R^2 /	.013/.023			0.052/			0.049/		
Conditional R^2				0.072			0.071		

Note Stars indicate level of significance with ^{ns} $p > .1$, ⁺ $p < .1$, ^{*} $p < .05$, ^{**} $p < .01$, and ^{***} $p < .001$.

Abbreviations: σ^2 , random effect variance; CI, confidence interval; ICC, intraclass correlation.

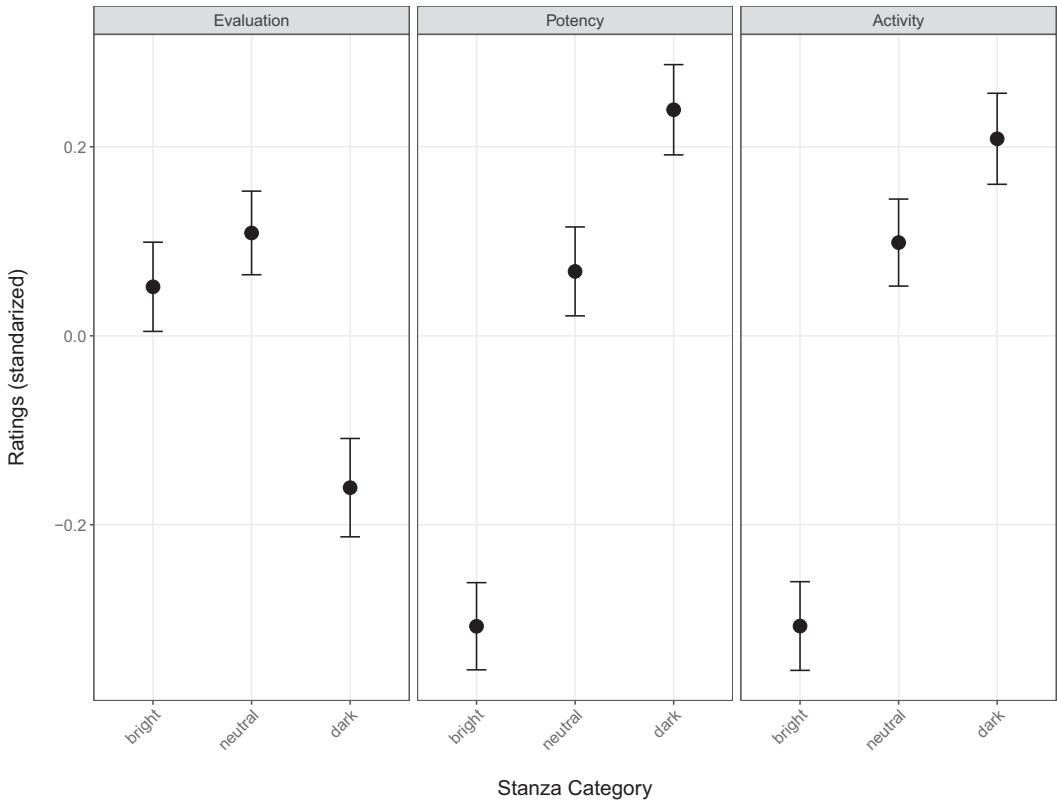


Fig. 3. Comparison between bright, neutral, and dark stanzas for each EPA dimension. The error bars indicate standard error of means.

stanzas. To test this, we conducted linear mixed models with the ratings per dimension as outcome variable, the average formant dispersion as predictor, and a random intercept per participant. The results suggest that the relation between the emotional tone of the stanzas and their formant dispersion is highly significant but weak (Evaluation: $B = 0.08$, $p < .01$, $R^2 = .006$; Potency: $B = -0.22$, $p < .001$, $R^2 = .051$; Activity: $B = -0.20$, $p < .001$, $R^2 = .040$). That is, while our data corroborate the assumption of a statistically significant correlation between formant dispersion and emotional tone, the amount of variance explained by the correlation (R^2) is relatively small, indicating that acoustic characteristics of vowels are only one among many features that influence the subjectively perceived emotional tone (for detailed results, see Materials S7 and S8).

Finally, in view of our data, it is also noteworthy that the ratings for the items of the Potency dimension and those of the Activity dimension were almost identical. Thus, it seems that there is a close interrelation between these two dimensions or, more cautiously, that our rating items are not suited to distinguish them. To statistically test the impression of a relation between Potency and Activity, we correlated the average ratings for each poem between each pair of EPA dimensions (Fig. 5). The correlations were

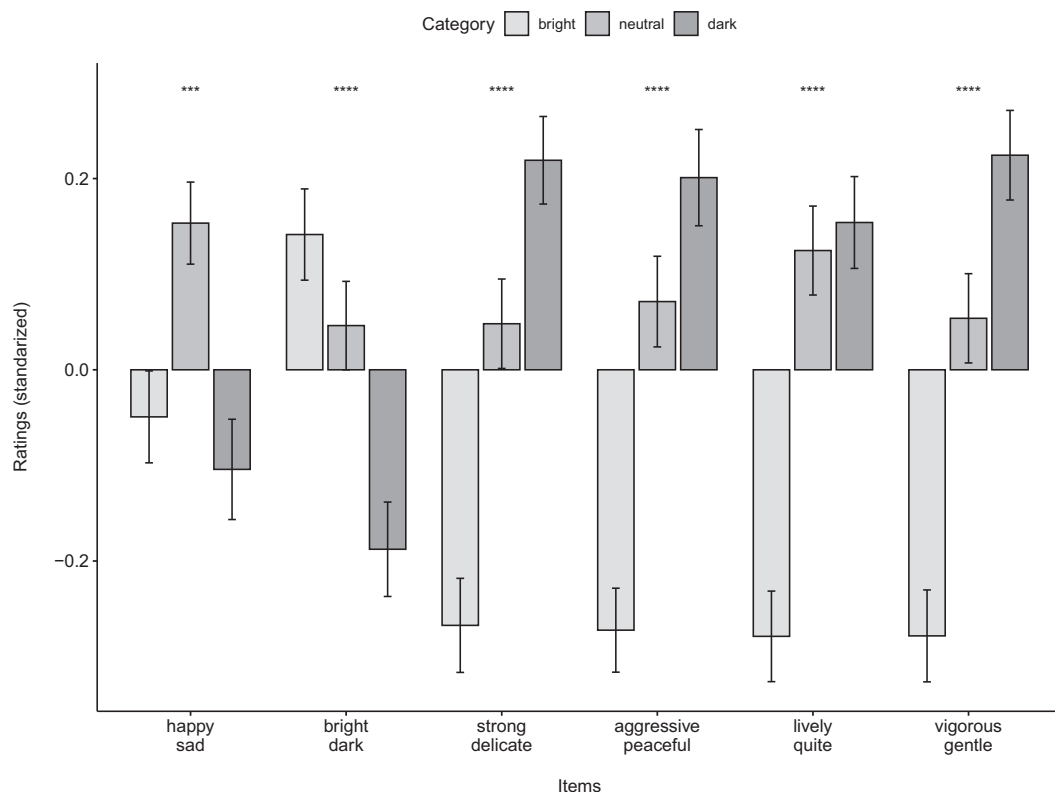


Fig. 4. Ratings per item for the three categories of stanzas (bright, neutral, dark). Positive values indicate ratings closer to the upper adjective of the respective antagonists. The error bars indicate standard error of means. Stars indicate level of significance with $***p < .001$ and $****p < .0001$.

significant for Evaluation–Potency ($r = -.42$, $p < .001$) and Potency–Activity ($r = .84$, $p < .001$) but not for Evaluation–Activity ($r = -.01$, $p > .1$). Moreover, as already indicated by Fig. 3, the correlation was considerably stronger for Potency–Activity than for Evaluation–Potency. We also conducted separate correlations for each category of stanzas (i.e., bright, neutral, dark). The results indicate that correlations between dimensions were by and large similar across the categories of the stanzas (for detailed results, see Material S9). The only noticeable difference is a highly significant correlation between Evaluation and Potency for neutral stanzas.

4. Discussion

We tested the predictive power of acoustic features of vowels to assess the emotional tone in poems. To this end, we calculated the average formant dispersion (i.e., the distance between the first two formants of a vowel) for 8,031 stanzas retrieved from a

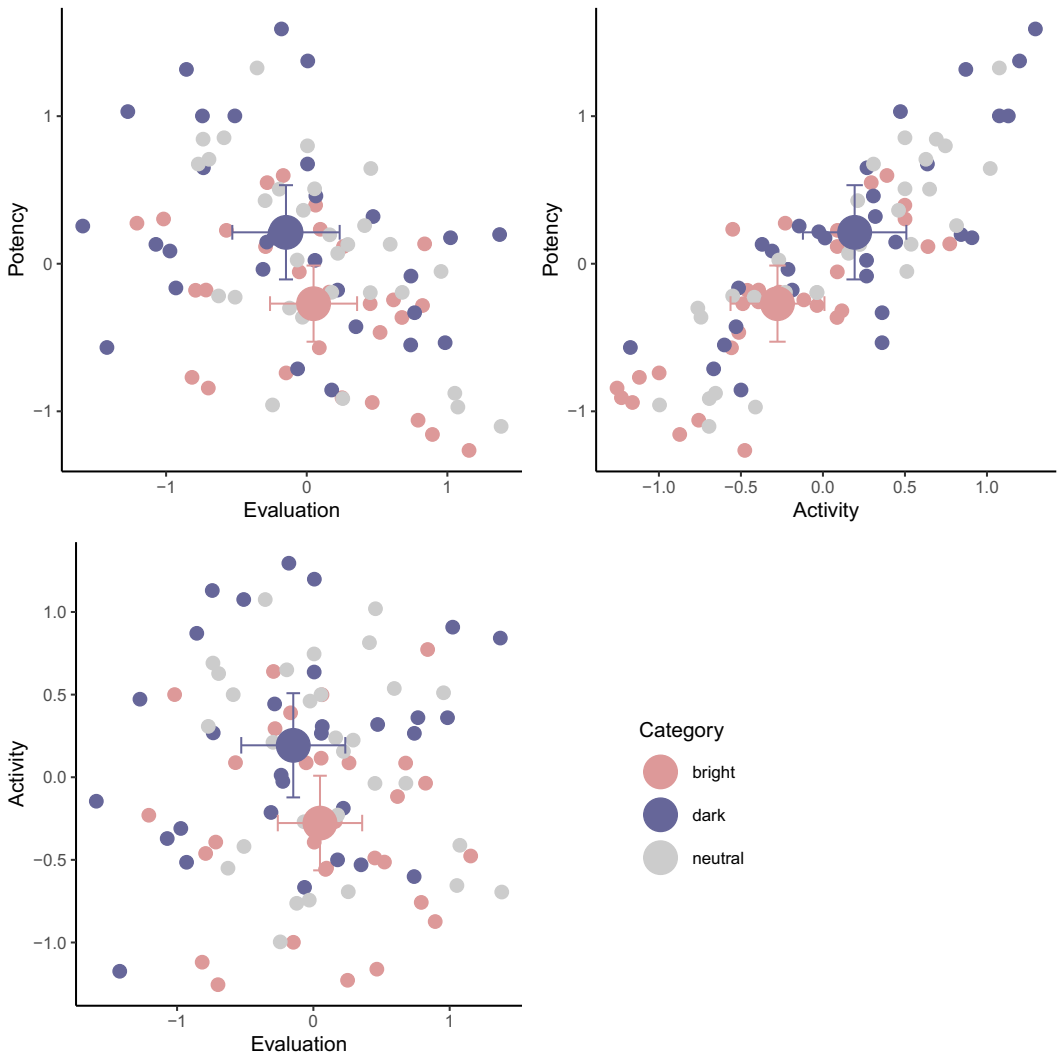


Fig. 5. Correlations between EPA dimensions for bright and dark stanzas. Small dots represent the average ratings for each stanza and bold dots those for the category. The error bars indicate standard deviations.

corpus of German poems. Participants assessed a random selection of 90 stanzas with either an extremely high formant dispersion (bright stanzas), an extremely low formant dispersion (dark stanzas), or an average formant dispersion (neutral stanzas) on six bipolar items comprising the three dimensions *Evaluation*, *Potency*, and *Activity*. Based on previous studies, we hypothesized that assessments of bright and dark stanzas would differ significantly on all three dimensions, with bright stanzas being rated higher on items referring to a positive *Evaluation* and a high level of *Activity*, whereas dark stanzas were expected to be rated higher on *Potency*.

Our results support this hypothesis for the Evaluation and the Potency dimensions of the EPA model. The assessment of the bright and dark stanzas indeed differed significantly on all items designed to measure these dimensions. What is more, except for the item *happy–sad*, the average rating for the neutral stanzas was always close to zero, whereas those of the two extremes (i.e., bright and dark stanzas) consistently diverged toward the opposite poles of the bipolar items. Thus, our data clearly suggest that acoustic characteristics of texts (passages) can be used as an indicator for the emotional tone of a text's content.

The results for the Activity dimension are not in line with our anticipations. In contrast to the results of previous studies, we found that dark stanzas (low formant dispersion) were mostly assessed as *lively* [lebhaft] and *vigorous* [energisch], whereas bright stanzas were assessed as *quiet* [ruhig] and *gentle* [sanft]. The strong correlation between the Potency and the Activity dimensions found in our data points to one possible explanation for this discrepancy between our results and those of previous studies. According to Fitch (1997), associations between formant dispersions and size may be rooted in the close relation between body size and length of the articulatory tract in mammals, including humans (Fitch & Giedd, 1999). The length of the articulatory tract determines the frequency of the formants of an individual's voice; accordingly, larger animals tend to have voices with a lower formant dispersion. For most animal species, size correlates with strength and strength with social dominance. Most likely for this reason, particularly male individuals purposefully lower the formant dispersion of their voice in order to appear larger and thereby to gain an advantage in the fight for resources (Charlton & Reby, 2016). Similar associations between formant dispersion and the notion of size, strength, and dominance have also been reported for humans (Evans, Neave, & Wakelin, 2006; Puts, Hodges, Cárdenas, & Gaulin, 2007; Watkins & Pisanski, 2016). Thus, formant dispersion is closely related to the position of an individual in the social hierarchy of an in-group. On a similar note, the behavior of dominant individuals is usually marked by more active, boastful, and aggressive behavior, whereas submissive behavior is likely to be more passive and self-derogatory as well as to avoid conflict (Holekamp & Strauss, 2016; Morton, 1977; Trower & Gilbert, 1989). Consequently, from an ethological perspective, there is a close relation between Potency (or Dominance) and Activity.

We, therefore, speculate that the relations between the acoustic features of vowels and the notions of activity might be derivative from the natural relation between formant dispersion and size. Our explanation for the discrepancy between our results and the results of previous studies is, therefore, as follows: First, we assume that there is a general tendency to implicitly associate large, strong, and heavy bodies with slow and clumsy movements. Second, participants who are asked to relate sound frequency to activeness or passiveness in a context-free research design might draw on this implied relationship between size and activeness and therefore end up assessing dark sounds as passive and bright sounds as active. In contrast, in our study, participants presumably assessed the stanzas based on their content rather than on their sound. Although we did not perform a detailed content analysis of the stanzas we used, it appears that knights—or, generally, warriors—appear relatively frequently in dark stanzas, whereas bright stanzas tend to deal

with love, lovesickness, the joy of love, and so forth. That is, the relation between dark vowels and Activity we observe in our data may be driven by the fact that the semantic concepts of knights and warfare are associated with a high degree of physical activity, irrespective of phonosemantic properties of the verbal material.

Our results also indicate that the relation between the emotional tone of a text and its acoustic features is not necessarily bipolar. According to our results, an extremely high average formant dispersion in a text is a good indicator for a low level of Potency (i.e., small and weak). In contrast, comparing the ratings for Potency between stanzas with a low average value for formant dispersion and neutral stanzas, the distance is relatively small (although statistically significant). Similarly, an extremely low formant dispersion turned out to be a good predictor for low ratings on items referring to Evaluation, whereas extremely high values for formant dispersion showed no significant effect.

These results are partly in line with previous findings. Studies that have compared relations between the lexical meaning of words and their phonetic characteristics across languages have shown that words and morphemes that refer to smallness have a greater than chance probability of the occurrence of high front vowels (Blasi et al., 2016; Ultan, 1978), indicating a close relationship between a high formant dispersion and a low level of Potency. In contrast, neither Ultan nor Blasi and colleagues reported evidence for a relation between the notion of high potency and the occurrence of dark-sounding vowels. Additionally, in a recent study, Westbury, Hollis, Sidhu, and Pexman (2018) found evidence suggesting that acoustic cues that are predictive for one pole of a dimension are not necessarily also a good predictor (with a reversed sign) for the opposite pole. Rather, it appears that tendencies to associate opposing acoustic features (e.g., high vs. low pitch) with opposing poles of semantic dimensions (e.g., small vs. large) often result from research designs introducing items as antonyms so that one pole is conceptualized as the negative of the other pole (e.g., *small* as *not large*). Therefore, we conclude that a relation of one acoustic feature (e.g., extremely low average formant dispersion) to a specific emotional tone (e.g., negative emotional valence) does not automatically imply that the two opposite poles (e.g., high formant dispersion and positive emotional valence) are likewise related to one another. In other words, the conclusion *e contrario* may simply not apply.

Yet another possible explanation lies in the very design of our study. It is, for example, thinkable that the relatively low values of bright stanzas for Evaluation result from the semantic focus of the items used. Note, for example, that the opposition between *happy* and *sad* also contains an element of Activity, as happiness mostly comes along with a high degree of activity and sadness with a low degree of activity. Consequently, it could be that bright stanzas simply referred to a positive-relaxed feeling that is not covered by the word “happy.” One problem with this explanation is that, for each dimension, neutral stanzas significantly differed with one pole of the extreme stanzas (e.g., between neutral and dark stanzas for Evaluation). If the lack of clear differences between neutral stanzas and bright stanzas for Evaluation, on the one hand, and dark stanzas for Activity, on the other hand, is due to deficient selection of items, then why does this only effect one extreme and not the other? As our data do not allow us to answer this question, we believe that further studies will be necessary to get a clearer picture.

Another limitation of this study is its preclusive focus on one specific phonetic feature, namely formant dispersion, to predict readers' assessment of the poems' emotional tone. The decision to emphasize this specific feature was based on a long tradition of research that provided sound evidence for associations between formant dispersion and concepts that refer to the dimensions of the Evaluation–Activity–Potency model. In other words, the driving question that motivated this study was whether a specific sound–meaning relation that has repeatedly been observed in studies from various academic disciplines can also be found in poetry. At the same time, results from recent studies suggest that there are also other features that proved to be statistically significant predictors for phonosemantic relations, such as articulatory characteristics of phonemes or the frequency of occurrence of individual phonemes (e.g., Louwerse & Qu, 2017; Nastase et al., 2007; Ohtake & Haryu, 2013). In fact, Monaghan and Fletcher (2019) reported that the frequency of individual phonemes outperforms phonetic features as a predictor for sound–meaning relations of nonwords. Thus, while our study reveals a statistically significant relation between the average formant dispersion of a text and its emotional tone, it does not allow drawing conclusions regarding the potential of formant dispersion as a predictor for emotional tone in relation to other features. As the formant dispersion of a vowel is causally linked to its articulatory characteristics, we cannot exclude the possibility that phonological features or vowel categories drive the results reported here. However, the aim of this study was to test whether findings reported in previous studies on sound–meaning relations also apply to texts. Our results suggest that this is the case. Future studies ought to more directly examine whether indeed formant dispersion or more abstract properties best explain the variance in the data—a comparison that is beyond the scope of this article.

Finally, we note that our results mainly reflect a distinction between a small group of stanzas with extreme values for a specific acoustic feature. Thus, it could well be that systematic relations between phonetic properties of texts and their content only come into play when the phonetic characteristics of a text exceed a certain threshold level. Evidence for the idea that sound–meaning relations in natural language only become salient when they strongly deviate from a norm has also been reported in previous studies. Aryani and colleagues, for example, found that phonetic features that are statistically overrepresented compared to their frequency of occurrence in everyday language can also predict the emotional tone of poems (Aryani, Kraxenberger, Ullrich, Jacobs, & Conrad, 2016; see also Aryani et al., 2013). Furthermore, similar evidence comes from a comparison between results reported by Auracher et al. (2010) and Kraxenberger and Menninghaus (2016). Both studies tested the relation between the average occurrence of certain phonemes (plosives and nasals) and the emotional tone of texts. However, whereas the first study did find significant interactions between sound and meaning for poems in four different languages, the data of the latter study could not replicate these results. One possible explanation for this discrepancy might lie in the research design: While in the study by Auracher et al. (2010) only poems with extreme values for the plosive–nasal ratio were used as stimuli, Kraxenberger and Menninghaus (2016) selected the poems used in their study on the basis of their emotional content.

Hence, it seems that phonosemantic relations in natural language only become salient for readers when phonetic features are foregrounded due to the unusually high frequency with which they occur in a text passage. We propose that congruencies between the acoustic properties and non-acoustic attributes of phonemes function as stylistic features, allowing authors to guide the attention of readers. However, as with any stylistic feature, authors are free to choose whether to use them or not. Thus, even though we found a highly significant correlation between formant dispersion and readers' assessment of the stanzas' emotional tone, our data also suggest that the relation between a text's acoustic characteristics and its emotional tone is far from being deterministic. Rather, for all three categories of stanzas used in our study (i.e., stanzas with extremely high, extremely low, and neutral formant dispersion), we obtained ratings for the three EPA dimensions that covered the full spectrum from low to high. This might also explain why some studies with relatively small corpora have not found any phonosemantic relations (e.g., Kraxenberger & Menninghaus, 2016; Miall, 2001). However, while the content of a text does not necessarily determine the relative occurrence of specific phonetic features, an unusually high occurrence of a specific phonetic characteristic seems to significantly increase the likelihood that a text expresses a certain emotion tone. Thus, we assume that a high occurrence of a phonetic feature is a considerably good predictor for the content of a text.

5. Future work

Many phonosemantic relations, such as those between acoustic frequency and size, have been found to be independent of specific languages. Our results therefore call for replications both in other languages and also, going beyond poems, in other text genres. It would, moreover, be interesting to investigate whether or not sound iconicity can be used as a cross-linguistic tool for automatic text analysis.

Acknowledgment

Open access funding enabled and organized by Projekt DEAL.

References

- Adams, F. M., & Osgood, C. E. (1973). A cross-cultural study of the affective meanings of color. *Journal of Cross-Cultural Psychology*, 4(2), 135–156. <https://doi.org/10.1177/002202217300400201>
- Adelman, J. S., Estes, Z., & Cossu, M. (2018). Emotional sound symbolism: Languages rapidly signal valence via phonemes. *Cognition*, 175, 122–130. <https://doi.org/10.1016/j.cognition.2018.02.007>
- Akita, K. (2009). A grammar of sound-symbolic words in Japanese: Theoretical approaches to iconic and lexical properties of mimetics (Doctoral dissertation). Retrieved from Kobe University Repository. (D1004724).

- Allott, R. (1995). Sound symbolism. In U. L. Figge (Ed.), *Language in the Würm glaciation* (pp. 15–38). Bochum: Brockmeyer.
- Aryani, A., Conrad, M., & Jacobs, A. M. (2013). Extracting salient sublexical units from written texts: "Emophon," a corpus-based approach to phonological iconicity. *Frontiers in Psychology*, 4, 654. <https://doi.org/10.3389/fpsyg.2013.00654>
- Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018). Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making. *PLoS One*, 13(6), e0198430. <https://doi.org/10.1371/journal.pone.0198430>
- Aryani, A., Hsu, C.-T., & Jacobs, A. M. (2018). The sound of words evokes affective brain responses. *Brain Sciences*, 8(6), 94. <https://doi.org/10.3390/brainsci8060094>
- Aryani, A., Kraxenberger, M., Ullrich, S., Jacobs, A., & Conrad, M. (2016). Measuring the basic affective tone of poems via phonological saliency and iconicity. *Psychology of Aesthetics, Creativity, and the Arts*, 10(2), 191–204. <https://doi.org/10.1037/aca0000033>
- Auracher, J. (2017). Sound iconicity of abstract concepts: Place of articulation is implicitly associated with abstract concepts of size and social dominance. *PLoS One*, 12(11), e0187196. <https://doi.org/10.1371/journal.pone.0187196>
- Auracher, J. (2020). Methodological issues in the study of phonetic symbolism (commentary to Tsur and Gafni). *Scientific Study of Literature*, 9(2), 230–239. <https://doi.org/10.1075/ssol.19012.aur>
- Auracher, J., Albers, S., Zhai, Y., Gareeva, G., & Stavniychuk, T. (2010). P is for happiness, N is for sadness: Universals in sound iconicity to detect emotions in poetry. *Discourse Processes*, 48(1), 1–25. <https://doi.org/10.1080/01638531003674894>
- Auracher, J., Scharinger, M., & Menninghaus, W. (2019). Contiguity-based sound iconicity: The meaning of words resonates with phonetic properties of their immediate verbal contexts. *PLoS One*, 14(5), e0216930. <https://doi.org/10.1371/journal.pone.0216930>
- Bakker, I., van der Voordt, T., Vink, P., & de Boon, J. (2014). Pleasure, arousal, dominance: Mehrabian and Russell revisited. *Current Psychology*, 33(3), 405–421. <https://doi.org/10.1007/s12144-014-9219-4>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Becker, J. A., & Fisher, S. K. (1988). Comparison of associations to vowel speech sounds by English and Spanish speakers. *The American Journal of Psychology*, 101(1), 51–57.
- Bentley, M., & Varon, E. J. (1933). An accessory study of "Phonetic Symbolism". *The American Journal of Psychology*, 45(1), 76–86. <https://doi.org/10.2307/1414187>
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences of the United States of America*, 113(39), 10818–10823. <https://doi.org/10.1073/pnas.1605782113>
- Charlton, B., & Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications*, 7, 12739. <https://doi.org/10.1038/ncomms12739>
- Coulter, K. S. (2009). The effects of phonetic symbolism on comparative price perceptions. In A. L. McGill & S. Shavitt (Eds.), *Advances in consumer research* (Vol. 36, pp. 986–987). Duluth, MN: Association for Consumer Research.
- Coulter, K. S., & Coulter, R. A. (2010). Small sounds, big deals: Phonetic symbolism effects in pricing. *Journal of Consumer Research*, 37(2), 315–328. <https://doi.org/10.1086/651241>
- Crawford, L. E. (2009). Conceptual metaphors of affect. *Emotion Review*, 1(2), 129–139. <https://doi.org/10.1177/1754073908100438>
- Cuskley, C., Dingemans, M., Kirby, S., & van Leeuwen, T. M. (2019). Cross-modal associations and synesthesia: Categorical perception and structure in vowel–color mappings in a large online sample. *Behavior Research Methods*, 51(4), 1651–1675. <https://doi.org/10.3758/s13428-019-01203-7>
- Dautriche, I., Mahowald, K., Gibson, E., & Piantadosi, S. T. (2017). Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cognitive Science*, 41(8), 2149–2169. <https://doi.org/10.1111/cogs.12453>

- Davis, C. P., Morrow, H. M., & Lupyan, G. (2019). What does a horgous look like? Nonsense words elicit meaningful drawings. *Cognitive Science*, 43(10), e12791. <https://doi.org/10.1111/cogs.12791>
- Dingemans, 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. <https://doi.org/10.1016/j.tics.2015.07.013>
- Duduciuc, A. C. (2015). Advertising brands by means of sound symbolism: The influence of vowels on perceived brand characteristics. *Studies and Scientific Researches: Economics*, 21, 112–119. <https://doi.org/10.29358/scceo.v0i21.304>
- Ertel, S. (1965). Der Lautcharakter Künstlicher Lautgebilde. *Psychologische Forschung*, 28, 491–518. <https://doi.org/10.1007/BF00421741>
- Evans, S., Neave, N., & Wakelin, D. (2006). Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice. *Biological Psychology*, 72(2), 160–163. <https://doi.org/10.1016/j.biopsycho.2005.09.003>
- Fischer-Jørgensen, E. (1968). Perceptual dimensions of vowels. *STUF: Language Typology and Universals*, 21(1–6), 94–98. <https://doi.org/10.1524/stuf.1968.21.16.94>
- Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, 102(2), 1213–1222. <https://doi.org/10.1121/1.421048>
- Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America*, 106(3), 1511–1522. <https://doi.org/10.1121/1.427148>
- Folkins, C., & Lenrow, P. B. (1966). An investigation of the expressive values of graphemes. *Psychological Record*, 16, 193–200. <https://doi.org/10.1007/BF03393659>
- Fónagy, I. (1961). Communication in poetry. *WORD*, 17(2), 194–218. <https://doi.org/10.1080/00437956.1961.11659754>
- Fónagy, I. (1963). *Die Metaphern in der Phonetik: Ein Beitrag zur Entwicklungsgeschichte des Wissenschaftlichen Denkens [Metaphors in phonetics: A contribution to the history of the development of scientific thought]*. The Hague, Netherlands: Mouton.
- Goudbeek, M. B., Goldman, J. P., & Scherer, K. R. (2009). Emotion dimensions and formant position. In M. Uther, R. Moore, & S. Cox (Eds.), *Proceedings of Interspeech 2009: 10th Annual Conference of the International Speech Communication Association* (pp. 1575–1578). https://www.isca-speech.org/archive/interspeech_2009/i09_1575.html
- Greenberg, J. H., & Jenkins, J. J. (1966). Studies in the psychological correlates of the sound system of American English. *WORD*, 22(1–3), 207–242. <https://doi.org/10.1080/00437956.1966.11435451>
- Heise, D. R. (1966). Sound–meaning correlations among 1,000 English words. *Language and Speech*, 9(1), 14–27. <https://doi.org/10.1177/002383096600900102>
- Herold, D. S., Nygaard, L. C., & Namy, L. L. (2012). Say it like you mean it: Mothers’ use of prosody to convey word meaning. *Language and Speech*, 55(3), 423–436. <https://doi.org/10.1177/0023830911422212>
- Hinton, L., Nichols, J., & Ohala, J. J. (1994). Introduction: Sound-symbolic processes. In J. Nichols, J. J. Ohala, & L. Hinton (Eds.), *Sound symbolism* (1st ed., pp. 1–14). Cambridge, England: Cambridge University Press.
- Holekamp, K. E., & Strauss, E. D. (2016). Aggression and dominance: An interdisciplinary overview. *Current Opinion in Behavioral Sciences*, 12, 44–51. <https://doi.org/10.1016/j.cobeha.2016.08.005>
- Hoshi, H., Kwon, N., Akita, K., & Auracher, J. (2019). Semantic associations dominate over perceptual associations in vowel–size iconicity. *I-Perception*, 10(4), 2041669519861981. <https://doi.org/10.1177/2041669519861981>
- Humboldt, W. V. (1836). Über die Verschiedenheit des menschlichen Sprachbaus und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts [The heterogeneity of language and its influence on the intellectual development of mankind]. Available at: <https://ia800203.us.archive.org/26/items/berdieverschied00humbgoog/berdieverschied00humbgoog.pdf> Accessed Sept. 22, 2020

- Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 369(1651), 20130298. <https://doi.org/10.1098/rstb.2013.0298>
- Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109(1), 54–65. <https://doi.org/10.1016/j.cognition.2008.07.015>
- Jakobson, R., & Waugh, L. R. (2002). *The sound shape of language*. Berlin, Germany: De Gruyter. <https://doi.org/10.1515/9783110889451>. Available at: <http://www.degruyter.com/doi/book/10.1515/9783110889451> Accessed Sept. 22, 2020
- Jespersen, O. (1922). Language: Its nature, development, and origin. Chapter 20: Sound symbolism. Available at: <https://www.gutenberg.org/files/53038/53038-h/53038-h.htm> Accessed Sept. 22, 2020
- Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science*, 35(3), 575–586. <https://doi.org/10.1111/j.1551-6709.2010.01169.x>
- Kjetsaa, G. (1974). Lomonosov's sound characteristics. *Scando-Slavica*, 20(1), 77–93.
- Klink, R. R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters*, 11(1), 5–20. <https://doi.org/10.1023/A:1008184423824>
- Klink, R. R., & Wu, L. (2014). The role of position, type, and combination of sound symbolism imbeds in brand names. *Marketing Letters*, 25(1), 13–24. <https://doi.org/10.1007/s11002-013-9236-3>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting Intraclass Correlation Coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kraxenberger, M., & Menninghaus, W. (2016). Mimological reveries? Disconfirming the hypothesis of phono-emotional iconicity in poetry. *Frontiers in Psychology*, 7, 1779. <https://doi.org/10.3389/fpsyg.2016.01779>
- Ladefoged, P., & Disner, S. F. (2012). *Vowels and consonants* (3rd ed.). Chichester, England: Wiley-Blackwell. Available at: <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10528115>
- Laukka, P., Juslin, P. N., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion*, 19(5), 633–653. <https://doi.org/10.1080/02699930441000445>
- Leibniz, G. W. V. (1765) *Nouveaux essais sur l'entendement humain* [New essay on human understanding]. Available at: https://pdfs.semanticscholar.org/adaa/bc82224211ecb13347227d4b4d1c4bbe240f.pdf?_ga=2.216490838.1762961350.1585531133-1893496760.1558242201 Accessed Sept. 22, 2020
- Locke, J. (1690). An essay concerning human understanding. Book III: Words. Available at: <https://www.ebrary.com/assets/pdfs/locke1690book3.pdf> Accessed Sept. 22, 2020
- Lockwood, G., & Dingemans, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, 6, 1246. <https://doi.org/10.3389/fpsyg.2015.01246>
- Lockwood, G., Dingemans, M., & Hagoort, P. (2016). Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(8), 1274–1281. <https://doi.org/10.1037/xlm0000235>
- Lomonosov, M. (1748). *Kratkoe rukovodstvo k krasnorechiiu* [A short manual to rhetoric]. St. Petersburg: Imperatorskaya Akademia Nauk.
- Louwerse, M., & Qu, Z. (2017). Estimating valence from the sound of a word: Computational, experimental, and cross-linguistic evidence. *Psychonomic Bulletin & Review*, 24, 849–855. <https://doi.org/10.3758/s13423-016-1142-2>
- Lowrey, T. M., & Shrum, L. J. (2007). Phonetic symbolism and brand name preference. *Journal of Consumer Research*, 34(3), 406–414. <https://doi.org/10.1086/518530>
- Lüdtke, D. (2017). sjPlot: Data visualization for statistics in social science; R package version 2. Available at: <https://zenodo.org/record/2400856#.Xbf4i1UzZYE> Accessed Sept. 22, 2020
- Magnus, M. (2001). What is in a word? Studies of phonosemantics (doctoral dissertation). Trondheim, Norway: Norwegian University of Science and Technology. Available at: https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/243679/123799_FULLTEXT01.pdf Accessed Sept. 22, 2020

- Markel, N. N., & Hamp, E. P. (1960). Connotative meanings of certain phoneme sequences. *Studies in Linguistics*, 14(3–4), 47–61.
- Marks, L. E. (1978). *The unity of senses. Interrelations among the modalities*. New York: Academic Press.
- Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. Cambridge, MA: MIT Press.
- Meier, B. P., Robinson, M. D., Crawford, L. E., & Ahlvers, W. J. (2007). When "light" and "dark" thoughts become light and dark responses: Affect biases brightness judgments. *Emotion*, 7(2), 366–376. <https://doi.org/10.1037/1528-3542.7.2.366>
- Miall, D. S. (2001). Sounds of contrast: An empirical approach to phonemic iconicity. *Poetics*, 29(1), 55–70. [https://doi.org/10.1016/S0304-422X\(00\)00025-5](https://doi.org/10.1016/S0304-422X(00)00025-5)
- Miron, M. S. (1961). A crosslinguistic investigation of phonetic symbolism. *The Journal of Abnormal and Social Psychology*, 62(3), 623–630. <https://doi.org/10.1037/h0045212>
- Monaghan, P., & Fletcher, M. (2019). Do sound symbolism effects for written words relate to individual phonemes or to phoneme features? *Language and Cognition*, 11(2), 235–255. <https://doi.org/10.1017/langcog.2019.20>
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 369(1651), 20130299. <https://doi.org/10.1098/rstb.2013.0299>
- Moos, A., Smith, R., Miller, S. R., & Simmons, D. R. (2014). Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder. *I-Perception*, 5(2), 132–142. <https://doi.org/10.1068/i0626>
- Mortensen, D. R., Dalmia, S., & Littell, P. (2018). Epitran: Precision G2P for many languages. In N. Calzolari (Ed.), *Lrec 2018: Eleventh International Conference on Language Resources and Evaluation: May 7–12, 2018, Miyazaki, Japan* (pp. 2710–2714). Paris, France: European Language Resources Association.
- Morton, E. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, 111(981), 855–869. www.jstor.org/stable/2460385.
- Myers-Schulz, B., Pujara, M., Wolf, R. C., & Koehnigs, M. (2013). Inherent emotional quality of human speech sounds. *Cognition & Emotion*, 27(6), 1105–1113. <https://doi.org/10.1080/02699931.2012.754739>
- Nastase, V., Sokolova, M., & Shirabad, J. S. (2007). Do happy words sound happy?: A study of the relation between form and meaning for English words expressing emotions. In G. Angelova, R. Mitkov, & N. Nicolov (Eds.), *Recent advances in natural language processing: Proceedings [RANLP-2007]: Borovets, Bulgaria, 27–29 September 2007* (pp. 404–410). Shoumen, Bulgaria: Incoma.
- Newman, S. S. (1933). Further experiments in phonetic symbolism. *The American Journal of Psychology*, 45(1), 53–75. <https://doi.org/10.2307/1414186>.
- Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, 28(1), 225–252. <https://doi.org/10.1146/annurev.anthro.28.1.225>
- Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, 112(1), 181–186. <https://doi.org/10.1016/j.cognition.2009.04.001>
- Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, 33(1), 127–146. <https://doi.org/10.1111/j.1551-6709.2008.01007.x>
- Ohala, J. J. (1994). The frequency code underlies the sound-symbolic use of voice pitch. In J. Nichols, J. J. Ohala, & L. Hinton (Eds.), *Sound symbolism* (1st ed., pp. 325–347). Cambridge, England: Cambridge University Press.
- Ohtake, Y., & Haryu, E. (2013). Investigation of the process underpinning vowel–size correspondence. *Japanese Psychological Research*, 55(4), 390–399. <https://doi.org/10.1111/jpr.12029>
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana-Champaign, IL: University of Illinois Press.
- O’Shaughnessy, D. (1999). *Speech communications: Human and machine* (2nd ed.). New York: IEEE Press. <https://doi.org/10.1109/9780470546475>

- Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, *111* (16), 6104–6108. www.pnas.org/cgi/doi/10.1073/pnas.1322705111
- Parise, C. V., & Spence, C. (2013). Audiovisual cross-modal correspondences in the general population. In J. Simner & E. M. Hubbard (Eds.), *The Oxford handbook of synesthesia* (pp. 790–815). Oxford: Oxford University Press.
- Park, J. W., & Osera, S. (2008). The effect of brand sound on consumers' brand evaluation in Japan. In C. R. Acevedo, J. M. C. Hernandez, & T. M. Lowrey (Eds.), *Latin American advances in consumer research* (Vol. 2, pp. 188–189). Duluth, MN: Association for Consumer Research.
- Pätzold, M., & Simpson, A. P. (1997). Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. In A. P. Simpson, K. J. Kohler, & T. Rettstadt (Eds.), *AIPUK: Vol. 32 The Kiel Corpus of Read/Spontaneous Speech: Acoustic data base, processing tools and analysis results* (pp. 215–247). Kiel, Germany: Institut für Phonetik und digitale Sprachverarbeitung der Universität Kiel.
- Perlman, M., Clark, N., & Johansson Falck, M. (2015). Iconic prosody in story reading. *Cognitive Science*, *39*(6), 1348–1368. <https://doi.org/10.1111/cogs.12190>
- Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*(227), 1–15. <https://doi.org/10.3389/fpsyg.2010.00227>
- Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS One*, *10*(9), e0137147. <https://doi.org/10.1371/journal.pone.0137147>
- Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science*, *21*(3), e12572. <https://doi.org/10.1111/desc.12572>
- Plato (~360 BC). *Cratylus* (trans. Benjamin Jowett). Available at: <http://classics.mit.edu/Plato/cratylus.html> Accessed Sept. 22, 2020
- Pogacar, R., Plant, E., Rosulek, L. F., & Kouril, M. (2015). Sounds good: Phonetic sound patterns in top brand names. *Marketing Letters*, *26*(4), 549–563. <https://doi.org/10.1007/s11002-014-9288-z>
- Puts, D. A., Hodges, C. R., Cárdenas, R. A., & Gaulin, S. J. C. (2007). Men's voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340–344. <https://doi.org/10.1016/j.evolhumbehav.2007.05.002>
- R Core Team (2019). *R: A language and environment for statistical computing*. Available at: <https://www.R-project.org/> Accessed Sept. 22, 2020
- Reay, I. E. (1994). Sound symbolism. In R. E. Asher (Ed.), *The encyclopedia of language and linguistics* (1st ed., pp. 4064–4070). Oxford, England: Pergamon Press.
- Rummer, R., Schweppe, J., Schlegelmilch, R., & Grice, M. (2014). Mood is linked to vowel type: the role of articulatory movements. *Emotion*, *14*(2), 246–250. <https://doi.org/10.1037/a0035752>
- Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology*, *12*(3), 225–239. <https://doi.org/10.1037/h0070931>
- Schmidtke, D. S., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in Psychology*, *5*, 80. <https://doi.org/10.3389/fpsyg.2014.00080>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. <https://doi.org/10.1037//0033-2909.86.2.420>
- Shrum, L. J., Lowrey, T. M., Luna, D., Lerman, D. B., & Liu, M. (2012). Sound symbolism effects across languages: Implications for global brand names. *International Journal of Research in Marketing*, *29*(3), 275–279. <https://doi.org/10.1016/j.ijresmar.2012.03.002>
- Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*, *25*(5), 1619–1643. <https://doi.org/10.3758/s13423-017-1361-1>
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, and Psychophysics*, *73*, 971–995. <https://doi.org/10.3758/s13414-010-0073-7>

- Tarte, R. D. (1982). The relationship between monosyllables and pure tones: An investigation of phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior*, 21(3), 352–360. [https://doi.org/10.1016/S0022-5371\(82\)90670-3](https://doi.org/10.1016/S0022-5371(82)90670-3)
- Trnka, R., Lačev, A., Balcar, K., Kuška, M., & Tavel, P. (2016). Modeling semantic emotion space using a 3D hypercube-projection: An innovative analytical approach for the psychology of emotions. *Frontiers in Psychology*, 7, 522. <https://www.frontiersin.org/article/10.3389/fpsyg.2016.00522>
- Trower, P., & Gilbert, P. (1989). New theoretical conceptions of social anxiety and social phobia. *Clinical Psychology Review*, 9(1), 19–35. [https://doi.org/10.1016/0272-7358\(89\)90044-5](https://doi.org/10.1016/0272-7358(89)90044-5).
- Tsur, R. (2006). Size–sound symbolism revisited. *Journal of Pragmatics*, 38(6), 905–924. <https://doi.org/10.1016/j.pragma.2005.12.002>
- Tsur, R., & Gafni, C. (2020). Methodological issues in the study of phonetic symbolism. *Scientific Study of Literature*, 9(2), 195–229. <https://doi.org/10.1075/ssol.19001.tsu>
- Ullian, R. (1978). Size–sound symbolism. In J. H. Greenberg (Ed.), *Universals of human language: Phonology* (pp. 525–568). Stanford, CA: Stanford University Press.
- Vlasenko, B., Prylipko, D., Philippou-Hübner, D., & Wendemuth, A. (2011). Vowels formants analysis allows straightforward detection of high arousal acted and spontaneous emotions. Proceedings of Interspeech 2011: 12th Annual Conference of the International Speech Communication Association (pp. 1577–1580). [https://www.isca-speech.org/archive/interspeech_2011/Accessed Sept. 22, 2020](https://www.isca-speech.org/archive/interspeech_2011/Accessed%20Sept.%2022,%202020)
- Watkins, C. D., & Pisanski, K. (2016). Vocal indicators of dominance. In T. K. Shackelford & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of evolutionary psychological science* (pp. 1–6). New York: Springer.
- Wescott, R. W. (1971). Linguistic iconism. *Language*, 47(2), 416–428. <https://doi.org/10.2307/412089>
- Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2018). Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. *Journal of Memory and Language*, 99, 122–150. <https://doi.org/10.1016/j.jml.2017.09.006>
- Whissell, C. (1999). Phonosymbolism and the emotional nature of sounds: Evidence of the preferential use of particular phonemes in texts of differing emotional tone. *Perceptual and Motor Skills*, 89(5), 19–48. <https://doi.org/10.2466/PMS.89.5.19-48>
- Whissell, C. (2003). The emotional symbolism of two English e-sounds: /a/ as in “cheap” is pleasant and /i/ as in “chip” active. *Perceptual and Motor Skills*, 96(1), 149. <https://doi.org/10.2466/PMS.96.1.149-165>
- Whissell, C. (2004a). Poetic emotion and poetic style: The 100 poems most frequently included in anthologies and the work of Emily Dickinson. *Empirical Studies of the Arts*, 22(1), 55–75. <https://doi.org/10.2190/FWGA-M9DB-P9D4-11X6>
- Whissell, C. (2004b). “The sound must seem an echo to the sense”: Pope’s use of sound to convey meaning in his translation of Homer’s Iliad. *Perceptual and Motor Skills*, 98(3, Pt 1), 859–864. <https://doi.org/10.2466/pms.98.3.859-864>
- Whissell, C. (2011). Sound and emotion in Milton’s paradise lost. *Perceptual and Motor Skills*, 113(1), 257–267. <https://doi.org/10.2466/04.21.28.PMS.113.4.257-267>
- Whissell, C. (2016). An episodic plot and an unhappy ending: What emotional sound patterns reveal about the structure of Coleridge’s Rime of the Ancient Mariner. *Psychological Reports*, 118(1), 266–276. <https://doi.org/10.1177/0033294115626824>
- Whissell, C. (2017). Sound symbolism in Shakespeare’s sonnets: Evidence of dramatic tension in the interplay of harsh and gentle sounds. *English Language and Literature Studies*, 7(4), 1. <https://doi.org/10.5539/ells.v7n4p1>
- Whissell, C. (2018). Emotional sound symbolism and the volta in Shakespearean and Petrarchan sonnets. *English Language and Literature Studies*, 8(1), 1. <https://doi.org/10.5539/ells.v8n1p1>
- Wrembel, M., & Rataj, K. (2008). “Sounds like a rainbow”: Sound–colour mappings in vowel perception. In A. Botinis (Ed.), Proceedings of the 2nd ISCA Workshop on Experimental Linguistics: ExLing 2 (pp. 237–240). Athens, Greece: University of Athens.

- Yorkston, E., & Menon, G. (2004). A sound idea: Phonetic effects of brand names on consumer judgments. *Journal of Consumer Research*, 31(1), 43–51. <https://doi.org/10.1086/383422>
- Yoshida, H. (2012). A cross-linguistic study of sound-symbolism in children’s verb learning. *Journal of Cognition and Development*, 13(2), 232–265. <https://doi.org/10.1080/15248372.2011.573515>

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Material S1. Studies on sound–meaning relations of vowels with non-acoustic concepts.

Material S2. Formant frequencies in Mel for German vowels (Pätzold and Simpson, 1997).

Material S3. Selected stanzas with 30 stanzas per category (bright, neutral, dark). "Stanza ID" corresponds to the column "Stanza ID" in the data (S5).

Material S4. Instructions as used in the experiment (English translation).

Material S5. Data with “ID”: consecutive numbering of participants; “Run”: experimental run; “Stanza ID”: identification number of stanza as used in Supplementary materials S3; “Category”: category of stanza (i.e.: “bright”, “dark”, “neutral”); “mean_dF”: averaged Formant Dispersion of stanza; “happy” to “vigorous”: ratings of stanza; “Evaluation” – “Activity”: mean values of ratings per EPA-dimension.

Material S6. Post hoc Tukey test for multiple comparison of means (with 95% CI).

Material S7. Coefficients (Estimates) for *formant dispersion* for each EPA dimension (Model: $\text{Dim} \sim \text{FD} + 1 | \text{Part}$).

Material S8. Regression lines for ratings in dependency of formant dispersion separated for EPA-dimensions.

Material S9. Correlations of average ratings for each poem between EPA dimensions.

Appendix

Table A1

Ratings of stanzas by category, separately for the three EPA dimensions ($N = 430$)

	Evaluation			Potency			Activity		
	<i>M</i>	<i>SE</i>	CI	<i>M</i>	<i>SE</i>	CI	<i>M</i>	<i>SE</i>	CI
Bright	0.052	0.047	0.093	-0.307	0.046	0.091	-0.307	0.047	0.092
Neutral	0.109	0.044	0.087	0.068	0.047	0.092	0.099	0.046	0.090
Dark	-0.161	0.052	0.102	0.239	0.048	0.094	0.208	0.048	0.095

Note. CI, confidence interval; *M*, mean; *SE*, standard error.

Table A2

Mean values for each item, separately for the three categories ($N = 430$)

	Bright			Neutral			Dark			ANOVA	
	<i>M</i>	<i>SE</i>	CI	<i>M</i>	<i>SE</i>	CI	<i>M</i>	<i>SE</i>	CI	<i>df</i>	<i>F</i>
Happy	-0.049	0.048	0.094	0.153	0.043	0.084	-0.104	0.052	0.103	2, 1,287	7.99***
Bright	0.141	0.048	0.094	0.046	0.046	0.091	-0.188	0.049	0.097	2, 1,287	12.56***
Strong	-0.267	0.049	0.097	0.048	0.047	0.092	0.219	0.046	0.090	2, 1,287	27.25***
Aggressive	-0.272	0.044	0.086	0.071	0.047	0.093	0.201	0.050	0.099	2, 1,287	26.75***
Lively	-0.279	0.047	0.093	0.125	0.047	0.091	0.154	0.048	0.094	2, 1,287	26.14***
Vigorous	-0.278	0.048	0.094	0.054	0.047	0.092	0.224	0.047	0.092	2, 1,287	29.34***

Note. Stars indicate the level of significance with *** $p < .001$.

Abbreviations: CI, confidence interval; *df*, degree of freedom; *M*, mean; *SE*, standard error.