



# Prosody-Based Sound-Emotion Associations in Poetry

Maria Kraxenberger<sup>1\*</sup>, Winfried Menninghaus<sup>1</sup>, Anna Roth<sup>2</sup> and Mathias Scharinger<sup>1,3</sup>

<sup>1</sup> Language and Literature, Max-Planck-Institut für empirische Ästhetik, Frankfurt, Germany, <sup>2</sup> Department for Psycholinguistics and Didactics of German, Goethe-Universität Frankfurt am Main, Frankfurt, Germany, <sup>3</sup> Phonetics Research Group, Department of German Linguistics, Center for Mind, Brain and Behavior, Philipps-Universität Marburg, Marburg, Germany

Conveying emotions in spoken poetry may be based on a poem's semantic content and/or on emotional prosody, i.e., on acoustic features above single speech sounds. However, hypotheses of more direct sound–emotion relations in poetry, such as those based on the frequency of occurrence of certain phonemes, have not withstood empirical (re)testing. Therefore, we investigated sound–emotion associations based on prosodic features as a potential alternative route for the, at least partially, non-semantic expression and perception of emotions in poetry. We first conducted a pre-study designed to validate relevant parameters of joy- and sadness-supporting prosody in the recitation, i.e. acoustic production, of poetry. The parameters obtained thereof guided the experimental modification of recordings of German joyful and sad poems such that for each poem, three prosodic variants were constructed: one with a joy-supporting prosody, one with a sadness-supporting prosody, and a neutral variant. In the subsequent experiment, native German speakers and participants with no command of German rated the joyfulness and sadness of these three variants. This design allowed us to investigate the role of emotional prosody, operationalized in terms of sound-emotion parameters, both in combination with and dissociated from semantic access to the emotional content of the poems. The findings from our pre-study showed that the emotional content of poems (based on pre-classifications into joyful and sad) indeed predicted the prosodic features pitch and articulation rate. The subsequent perception experiment revealed that cues provided by joyful and sad prosody specifically affect non-German-speaking listeners' emotion ratings of the poems. Thus, the present investigation lends support to the hypothesis of prosody-based iconic relations between perceived emotion and sound qualia. At the same time, our findings also highlight that semantic access substantially decreases the role of cross-language sound–emotion associations and indicate that non-German-speaking participants may also use phonetic and prosodic cues other than the ones that were targeted and manipulated here.

**Keywords:** emotional prosody, poetry, emotion perception, joy, sadness, articulation

## OPEN ACCESS

### Edited by:

Guillaume Thierry,  
Bangor University, United Kingdom

### Reviewed by:

Alexander Bergs,  
University of Osnabrück, Germany  
John E. Drury,  
Stony Brook University, United States

### \*Correspondence:

Maria Kraxenberger  
maria.kraxenberger@googlemail.com

### Specialty section:

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

**Received:** 23 February 2018

**Accepted:** 04 July 2018

**Published:** 25 July 2018

### Citation:

Kraxenberger M, Menninghaus W,  
Roth A and Scharinger M (2018)  
Prosody-Based Sound-Emotion  
Associations in Poetry.  
*Front. Psychol.* 9:1284.  
doi: 10.3389/fpsyg.2018.01284

## INTRODUCTION

Poetry is widely regarded as a genre in which semantic and formal components of language are interrelated in a particularly meaningful way (Jakobson, 1960; Menninghaus et al., 2017). According to literary theory, this results in a pronounced “palpableness” of form (Jakobson, 1960) and, most notably, of the sound patterns in poetry (Shklovsky, 1919/2012; Jakobson and Waugh, 1979/2002).

Theoretical reflections as well as empirical studies have repeatedly highlighted the relation between the sound of a poem and the perception of emotion by its readers (e.g., Valéry, 1958; Fónagy, 1961; Jakobson and Waugh, 1979/2002; Tsur, 1992; Whissell, 2002, 2011; Schrott and Jacobs, 2011; Aryani et al., 2016).

Several studies have advanced the idea of a relation between the frequency of occurrence of segmental units, i.e., of certain phonemes or phoneme classes, and emotion perception in poetry. For instance, one study reported that nasals are more frequent in Old Egyptian lamentations and ballads written by the German poet J. W. von Goethe, whereas plosives are more frequent in Old Egyptian and Goethean hymns (Albers, 2008). Similarly, another study reported that native speakers rated German, Chinese, Russian, and Ukrainian poems with a relatively high frequency of nasals as sad, and poems with a high frequency of plosives as happy/joyful (Auracher et al., 2010). A more recent study (Kraxenberger and Menninghaus, 2016a), however, could not replicate Auracher et al.'s (2010) findings and consequently questions a consistent frequency effect. Thus, it seems that the frequently assumed nexus between sound and emotion perception in poetry (e.g., Valéry, 1958; Fónagy, 1961; Tsur, 1992; Whissell, 2002, 2011) might not consistently be driven by the frequency of occurrence of certain phoneme classes.

We therefore tested the hypothesis that acoustic, suprasegmental features of emotional prosody might rather be a possible reason of this nexus. Furthermore, we compared poems that have previously been classified as either joyful or sad (Kraxenberger and Menninghaus, 2016b). This allowed us to examine both bottom-up information from emotional prosody and top-down information from semantic content. Based on previous research, we exclusively focused on the emotions joy and sadness, so called “basic emotions” that show up in facial, vocal and bodily expressive behaviors (Russell, 1980; Ekman, 1992; Jack et al., 2014). Both emotions will be referred to in their broadest sense, i.e., as representing their respective families of emotion (e.g., Scherer et al., 1991; Ekman, 1992). Thus, joy also encompasses emotions like happiness and elation, whereas sadness comprises the emotion of grief. Despite the fact that they are considered in this broad sense, joy and sadness still represent distinct, polar emotions with markedly different phenomenological qualia (Schmitz, 1969; Demmerling and Landweer, 2007; also see Kraxenberger and Menninghaus, 2016a,b). Also, and most importantly for this study, joy and sadness are commonly understood to be universally available (Ekman and Cordaro, 2011) and to serve a communicative function, for instance in terms of the emotional prosody of an utterance.

Emotional prosody is predominantly determined by the suprasegmental features of pitch, tempo, and intensity (Alter, 2002), and is conducive to conveying emotions (Paulmann, 2006). Empirical studies on acoustic cues in the expression of emotions have yielded largely consistent results (Scherer, 1986); this holds particularly for the expression and recognition of joy and sadness. Accordingly, several studies have shown that the emotions of joy and sadness can indeed be distinguished by

means of acoustic profiles that are dependent on suprasegmental parameters. In the case of joy, most studies have reported higher values for the mean F0 (pitch) and mean intensity, as well as a faster speech rate. The expression and recognition of sadness, by contrast, has been associated with lower values for these measures (cf. Kaiser, 1962; Van Bezooijen, 1984; Scherer, 1986, 2013; Banse and Scherer, 1996; Ververidis and Kotropoulos, 2006; Pell et al., 2009; Stolarski, 2015).

If joy and sadness can already be reliably detected by means of certain suprasegmental features in simple sentences and utterances, it is reasonable to conjecture that these parameters of emotional prosody should also yield a pronounced effect in poetry—a literary genre widely considered to be eminently emotional (e.g., Hegel, 1986; Winko, 2003; Lüdtke et al., 2014). However, despite the large amount of evidence supporting the idea of acoustic emotional profiles for smaller text units and utterances, the effects of suprasegmental features of emotional prosody on emotion perception in poetry recitation have rarely (if ever) been investigated to date. More importantly, most studies on emotional prosody focused on a bottom-up approach, i.e. they mainly considered acoustic features in relation to an emotional interpretation. The interaction of semantically based emotion content and emotion-supporting prosody, on the other hand, has not been studied before, particularly within poetry. Therefore, we set out and conducted a pre-study on the recitation, i.e., acoustic production of poems, as well as a perception experiment to test for prosody-based sound-emotion associations, each comparing sad and joyful poems.

The first study was designed to validate and establish quantitative measures for emotional prosody in native-speakers' poetry recitation of joyful and sad poems. To this end, participants were asked to read out aloud eight German poems. The suprasegmental features of emotional prosody obtained through this production study served as the basis for manipulating the acoustic presentation of joyful and sad poems in a subsequent perception experiment. This second study was conducted online; it aimed at testing whether and to what extent features of emotional prosody work only in conjunction with semantic understanding or can also influence emotion perception in the absence of semantic access. The study included two experimental conditions in which participants without and with knowledge of German (i.e., English and Japanese without German knowledge, as well as German participants, cf. page 6) were exposed to audio recordings of joyful and sad German poems that were systematically manipulated in terms of emotion-related suprasegmental parameters. In the first condition (Expressive Condition), we presented poems with emotionally expressive prosodies, i.e., a joy-supporting, and a sadness-supporting prosody. In the second condition (Neutral Condition), the same poems were presented with a “neutralized” emotional prosody. Targeting the prosodic features of the poems in this way and comparing participants with and without access to semantic-based emotional content allowed us to examine the relative weights of bottom-up (prosodic) and top-down (semantic) information during emotion perception in poetry.

## EXPERIMENT 1

### Materials and Methods

#### Participants

Twenty-three participants (15 females; ages ranging from 20 to 32 years,  $M = 25.0$ ,  $SD = 3.1$ ) took part in our study. Inclusion criteria for participation were German as a native language and full legal age. The participants were asked to recite and rate four joyful and four sad poems. All experimental procedures were ethically approved by the Ethics Council of the Max Planck Society and were undertaken with the informed consent of each participant.

#### Poems

We selected eight German poems written in the twentieth century. The poems had between 10 and 16 lines, and featured end rhymes (paired or cross rhymes), as well as a mainly iambic meter. In a previous study (Kraxenberger and Menninghaus, 2016b), four of these poems had been classified as joyful (Werner Bergengruen: *Sommersonett*; Paul Haller: *O leuchtender Septembertag*; Klavund: *Liebeslied, Dein Mund*; Joachim Ringelnatz: *Morgenwonne*), and four as sad (Yvan Goll: *Trauermarsch*; Ferdinand Hardekopf: *Spät*; Else Lasker-Schüler: *Dämmerung*; Jesse Thoor: *Die Zerwartung*). This classification was first based on matching the poems' main themes to phenomenological descriptions of joy and sadness (Schmitz, 1969; Demmerling and Landweer, 2007) and subsequently approved by an analysis of variance. Thus, the poems that were a priori classified as joyful turned out to be rated as significantly more joyful and as less sad than the poems classified as sad (see Kraxenberger and Menninghaus, 2016b).

#### Procedure

For recording their recitations, participants were seated in a sound-attenuated booth. The poems were presented on a screen in a randomized order. Participants were instructed to first read each poem silently. Subsequently, they were asked to rate *how joyful* (hereafter: Joy; German: *freudig*) and *how sad* (hereafter: Sadness; German: *traurig*) they perceived the respective poem to be, using seven-point items ranging from 1 (*not at all*) to 7 (*very much*)<sup>1</sup>. Next, participants were instructed to read the poem aloud. We used a directional headset microphone (DPA, d:fine) and the digital recording device Zoom H4n (sampling rate: 44.1 kHz, amplitude resolution: 16 bits) for recording. All participants read, rated, and recited all eight poems.

For calculating the acoustic measures of pitch and articulation rate, we first annotated each syllable of each recitation for all participants. We used WebMaus (Reichel, 2012; Reichel and Kisler, 2014) for syllabification and for the creation of text-based annotations. In a subsequent step, we controlled syllable boundaries and, if necessary, manually adjusted and corrected them to comply with phonetic conventions regarding phoneme segmentation. We then moved all boundaries to the nearest zero crossing and calculated the syllable rate. In line with Pfitzinger (2001), we analyzed the global net articulation rate, i.e., all

actually produced syllables, excluding pauses. Because pauses are known to influence the calculation of speech tempo (cf. Künzel, 1997; Trouvain et al., 2001; Jessen, 2007), and realized pauses during poetry recitation are strongly influenced by verse structure, we decided to exclude all pauses from our analyses of the articulation rate, along with repetitions, corrections, and nonverbal articulations (e.g., coughing). Although literature on acoustic emotion profiles often also report effects of intensity, we decided not to include this parameter in our experiments. This decision was based on the unpredictability of participants' individual tone and volume control, their technical equipment (e.g., sound cards, speaker system, etc.), and their default setting (or adjustment) of volume when participating in our subsequent online study.

#### Analysis

For each participant and poem, we divided the number of actually produced syllables by the total duration of all syllables, excluding pauses. We thereby obtained a value for the number syllables articulated per second for each participant and each recitation. Mean pitch was estimated using the fundamental frequency (F0) analysis in Praat (Boersma and Weenink, 2015). F0 was extracted from the voiced portions of the respective syllables based on an autocorrelation algorithm following the procedure proposed by Hirst (2011). We began with a floor value of 60 Hz and a ceiling value of 700 Hz. We excluded the first and last quartiles from our pitch data to compensate for outliers. Then we calculated F0 again, based on the new floor (minF0) and ceiling (maxF0) values. Potential variance between speakers was taken into account by applying random intercepts to our statistical analyses (see below)<sup>2</sup>.

#### Results

In a first step, we checked whether participants confirmed the preclassification of the poems as joyful or sad. Here we inspected the mean values of the Joy and Sadness ratings. The mean Joy ratings for the poems that were preclassified as joyful were all above the midpoint (4) of the seven-point item measuring Joy ( $M = 5.73$ ,  $SD = 1.19$ ). Likewise, the mean Sadness rating for the poems that were preclassified as sad were all well above the midpoint of the Sadness item ( $M = 5.57$ ,  $SD = 1.20$ ). Additionally, our analyses of variance showed significant differences between participants' ratings for Joy and Sadness of poems that were in line with our preclassification, i.e., poems preclassified as joyful were rated as more joyful and less sad than poems that were preclassified as sad (all  $p \leq 0.001$ , all  $\eta^2 \geq 0.77$ ).

To test whether the mean pitch of the participants' recitations was related to the perceived Joy and Sadness of the poems, we applied linear mixed models with random intercepts for poems and participants and the fixed effects of the participants' ratings for Joy and Sadness. Given a highly significant, negative correlation between the ratings for Joy and Sadness (Pearson Correlation, two-tailed,  $r = -0.91$ ,  $p \leq 0.001$ ), we included the ratings for Joy and Sadness in separate models. We also included

<sup>1</sup>Some additional items from the questionnaire were not considered in the analyses presented here.

<sup>2</sup>For a description of the packages and statistical software used, see pages 6–7.

participants' gender as an additional fixed effect (for an overview of gender effects in speech production see, for instance, Titze, 2000).

The results showed that participants' ratings of Joy were significantly related to the mean pitch at which they recited the poems. On average, an increase of joyfulness ratings by one unit on the 7 point scale corresponded to a 1.68 Hz increase in pitch ( $SE = 0.19$ ,  $t = 8.69$ ,  $p \leq 0.001$ ). Similarly, Sadness ratings significantly predicted the mean pitch of the recitations; here, an increase of sadness ratings by one unit corresponded to a 1.58 Hz decrease in pitch ( $SE = 0.23$ ,  $t = -6.97$ ,  $p \leq 0.001$ ). The same analysis was applied to test whether the articulation rate (i.e., number of syllables articulated per second) could likewise be predicted by the participants' ratings for Joy and Sadness. The results showed a relation of participants' Joy ratings and the tempo at which they recited the poems by trend ( $b = 0.02$ ,  $SE = 0.01$ ,  $t = 1.76$ ,  $p \leq 0.10$ ). The participants' ratings for Sadness showed a significant relation to the articulation rate ( $b = -0.03$ ,  $SE = 0.01$ ,  $t = -2.30$ ,  $p \leq 0.05$ ), indicating a decrease in tempo when ratings for Sadness increased.

The validation experiment thus provided us with mean pitch and mean rate changes per unit of Joy/Sadness rating. We subsequently used these values for our acoustic manipulations of joy-supporting and sadness-supporting prosody.

## EXPERIMENT 2

To test whether emotional prosody can be viewed as a pivotal factor *independent* of the semantic content of acoustically presented poems, and to investigate the role of semantic access with respect to participants' emotion ratings, we designed and conducted a cross-language study. Based on our validation of the suprasegmental features of emotional prosody during poetry recitation as reported in Experiment 1, we included two conditions in our survey (a so-called "Expressive Condition" and a "Neutral Condition"). In the Expressive Condition, we presented each joyful and sad poem with both joy- as well as with sadness-supporting prosodic cues, and hence in a semantic-matching and a semantic-mismatching variant with divergent prosody. In the Neutral Condition, we substantially reduced and hence "neutralized" the prosodic cues. Further, our design allowed us to test the effect of emotional prosody orthogonally to the participants' access to the poems' semantic content (German-speaking vs. non-German-speaking listeners), including, amongst others, syntactic cues, word frequency, or familiarity with German poetry. Our hypotheses and expectations were as follows:

- (a) The joy- and sadness-supporting prosodies in the Expressive Condition should substantially affect the emotion ratings of the non-German-speaking listeners who have no linguistic access to the poems' semantic content. More specifically, we expected a main effect of joy- and sadness supporting prosodies that would not interact with the preclassification of the poems as joyful or sad.
- (b) In the case of matching prosodic characteristics and emotional content (i.e., joyful poems presented with a

joy-supporting prosody), the prosodic distinction between sad and joyful poems should be particularly pronounced. Therefore, participants without semantic access should be able to distinguish between joyful and sad poems without understanding their content.

- (c) For the German-speaking listeners in the Expressive Condition, ratings should be strongly influenced by the preclassification of the poems as being joyful or sad in content. In an exploratory manner, we tested whether the potential effects of emotional prosody could be overridden by the semantic access of the German participants. Again, only a main effect of joy- and sadness supporting prosodies without interaction with the preclassification of the poems as joyful or sad would be considered as a pivotal factor.
- (d) In the absence of relevant prosodic cues in the Neutral Condition, the participants' ratings should depend only on the poems' semantic content; here, the emotion ratings of the German-speaking and non-German-speaking participants were expected to differ strongly from each other given that non-German speaking participants had no semantic access to the content of the poems.
- (e) German-speaking participants in the Neutral Condition were again expected to rate the poems in accordance to our preclassification of the poems as either joyful or sad.

In an exploratory manner, we tested in a first step across all conditions for differences in the ratings of English and Japanese participants (see page 6) that might be due different culture- or language-dependent stances and attitudes toward emotions, as well as due to the different language families English and Japanese are part of. We also tested for differences in the ratings of English and Japanese participants with and without experience in analyzing linguistic audio-material. The latter analyses were based on the assumption that such experience might influence the handling and evaluation of semantically incomprehensible audio material.

## Materials and Methods

### Recordings of Audio Stimuli

A professional actor recited the eight poems that were used in the validation study. During the recording session, the actor was instructed to read the poems in a rather restrained prosodic manner—i.e., without extreme modulations of voice, timbre, pitch, or tempo—and to avoid an excessive emphasis on metrical regularity and rhyme schemes.

To control for loudness differences between the individual stimuli, we decided to use a normalization based on the recommendation ITU-R BS.1770-3 (International Telecommunication Union, 2012). We used the open source tool R128GAIN (Belkner, 2014) to normalize each stimulus to a loudness value of  $-23\text{LUFS}$  (absolute measurement) according to R128<sup>3</sup>.

<sup>3</sup>R128 uses a loudness range to measure the variation of loudness on a macroscopic time scale in loudness units (LUs, relative measurement), where one LU corresponds to one Db. (For a more detailed description and the computation of the loudness range as a measurement of loudness levels, see Broadcasting Service (Sound) Series, 2012).

## Manipulation of Suprasegmental Features of Emotional Prosody

In order to assess the relative contribution that emotional prosody and the participants' access to the semantic content of the poems might make to the emotion ratings, we manipulated suprasegmental features of emotional prosody. All manipulations were done using Praat (Boersma and Weenink, 2015); they affected the mean pitch and tempo of the recordings. We also manipulated the poems' mean range of pitch to eliminate outliers with regard to single pitch points.

### Neutralized prosody

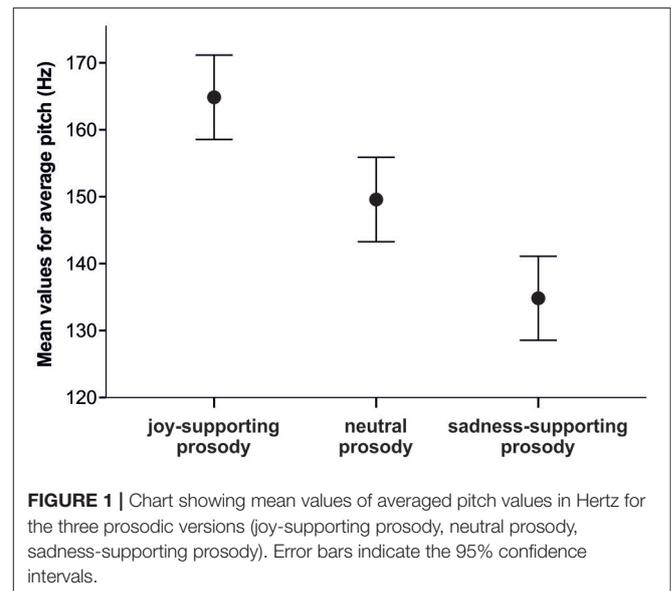
Before modifying the emotional prosody of our recordings in a joy- or sadness-supporting way, we first calculated the mean pitch across all eight recordings of the poems ( $M = 146.70$  Hz,  $SD = 7.30$ ,  $min = 135.82$  Hz,  $max = 158.24$ ). Then the mean pitch for all recorded poems was set to this mean. The mean pitch range was also adjusted accordingly ( $M = 353.43$  Hz,  $min = 75.017$  Hz,  $max = 428.445$  Hz). Thus, the manipulated recordings had the same mean values for pitch and pitch range but retained their original pitch contours. In this way, we produced recordings that had a natural-sounding human voice with controlled and less variable pitch values than the original recordings.

### Joy- and sadness-supporting prosodies

All subsequent pitch manipulations for our first experimental condition (Expressive Condition) were based on the neutral baseline obtained through the changes reported above. Since participants had shown an average pitch increase of 1.68 Hz per unit of joy rating and an average decrease of 1.58 Hz per unit of sadness rating in the validation study, we decided to change pitch values accordingly. To arrive at pitch levels associated with joy and sadness, we increased and decreased each pitch point by 15 Hz (about 10 times the changes per rating level within the validation study), respectively, starting from the neutral pitch levels (for an illustration of the mean pitches in each prosodic version, see **Figure 1**).

In addition to the pitch changes, we also changed the tempo of each recording of a poem with joy-supporting prosody by decreasing its length factor by 10%. For the emotionally expressive versions of the sad poems (i.e., sadness-supporting prosody), we applied the reverse procedure. That is, we lowered pitch points by 15 Hz and reduced the tempo by 10%. The versions that were to contain mismatching cues of emotional prosody were produced by applying the expressive emotional features of the sad poems to the joyful poems and vice versa. As a result, joyful poems were reduced in pitch and tempo, whereas sad poems were increased in pitch and tempo. Thus, we obtained two versions of each poem, one with an expressive, semantic-matching prosody (i.e., in line with our preclassification of the poems as either joyful or sad), and one with a reverse, mismatching emotional prosody (i.e., contrary to the preclassification).

In order to investigate the influence of semantic content independent of prosodic cues, we used the "neutralized" baseline versions as stimuli for a second condition (Neutral Condition). In this condition, acoustic cues should be rather



uninformative regarding emotional prosody, and we expected that German participants' ratings for Joy and Sadness would rely predominantly on their semantic access to the poems. Participants without knowledge of German, however, should have neither semantic nor prosodic cues guiding their emotion ratings.

## Participants

Native German, English, and Japanese speakers were recruited through postings on diverse mailing lists, through a distribution of postcards, and via social media sites. The inclusion criterion for participation in the study was being of full legal age. The participants did not receive monetary or other compensation. As for the validation study, the Ethics Council of the Max Planck Society ethically approved all experimental procedures. Prior to filling out the online questionnaire, participants provided their written informed consent. Non-German speakers were asked to indicate their German language skills using the level descriptions from the *Common European Framework of Reference for Languages* (CEFR). Participants with German-language skills at level B1 (Threshold) and above were excluded from the survey. All other participants were instructed regarding the procedure of the study.

Two hundred seventy-two participants completed the survey. Of these, 205 were German (137 indicated female gender and 9 participants did not specify their gender; ages ranged from 18 to 46 years,  $M = 32.6$  years,  $SD = 8.10$ ). Thirty-five complete data sets were obtained from participants who indicated that English was their mother tongue (21 participants indicated female gender and two participants did not specify their gender; ages ranged from 19 to 45 years,  $M = 33.8$  years,  $SD = 7.88$ ; 48.6 % indicated to have some kind of experience with analyzing linguistic audio material). Thirty-two complete data sets were obtained from native Japanese-speaking participants (22 participants were female and 10 were male; ages ranged from 19 to 45 years,

$M = 32.6$  years,  $SD = 9.70$ ; 9.4% indicated to have some kind of experience with analyzing linguistic audio material).

Among the native German-speaking group, 100 participants were assigned to the Expressive Condition and 105 participants to the Neutral Condition. Similarly, 16 native English speakers were assigned to the Expressive Condition and 19 to the Neutral Condition. Among the Japanese speakers, 18 were assigned to the Expressive Condition and 14 to the Neutral Condition. The assignment of participants to one of the two conditions was randomized.

### Online Questionnaire and Procedure

Ratings were collected using an online survey (in German, English, and Japanese)<sup>4</sup>. Participants were instructed to use headphones and to complete the survey alone without interruption. To become familiar with the rating procedure, participants listened to a practice poem and rated it. Afterwards, they were randomly assigned to one of the two conditions of the survey.

Participants assigned to the Expressive Condition were presented with two joyful and two sad poems with joy-supporting prosody, and two joyful and two sad poems with sadness-supporting prosody. We randomized the order and prosodic version of the poems per participant and made sure that participants would hear a poem only in one version. Participants in the neutral condition were exposed to all eight “neutralized” poems in randomized order. All audio files were stereo recordings with a 44.1 kHz sampling rate and 32 bit amplitude resolution.

After listening to a poem, participants were asked to indicate *how joyful* and *how sad* (German: *freudig/traurig*; Japanese: *楽しげな / 悲しい*)<sup>5</sup> they perceived the poems to be on a scale ranging from 1 (*not joyful/not sad at all*) to 7 (*very joyful/very sad*). Additionally, they rated the poems on a bipolar affect item ranging from 1 (*negative*) to 7 (*positive*) (for the distinction between emotion and affect, see Russell, 2003; Scherer, 2005). Ratings were gathered via seven-part visual analog scales (VASs; e.g., Aitken, 1969; Funke and Reips, 2012), allowing a fine-grained measurement of the participants’ responses. In addition to language skills and age, we also surveyed whether or not participants’ had any experience with the analysis of linguistic audio material. The average duration for survey completion was 17.36 min ( $SD = 6.13$ ,  $min = 10.63$ ,  $max = 49.6$ ).

In each condition, participants’ ratings for Joy, Sadness, and Negativity/Positivity correlated highly with one another (Pearson Correlation, two-tailed, all  $|r| \geq 0.82$ , all  $p \leq 0.001$ ). Thus, the bipolar item did not seem to provide substantially different information. For this reason, we exclusively report participants’ ratings for Joy and Sadness (hereafter: emotion ratings), in line with the preceding validation study.

<sup>4</sup>The survey was administered online between March 10 and July 25, 2016, using the software (Unipark QuestBack Globalpark AG., 2014). Some additional items from the questionnaire were not considered in the analyses presented here.

<sup>5</sup>We used the following phrases in the German and Japanese versions of our questionnaire: *Wie freudig ist das Gedicht? / Wie traurig ist das Gedicht?* この詩は楽しげですか? / この詩は悲しいですか? Please note that none of the participants commented negatively on the comprehensibility of our questionnaire.

## RESULTS

### Ratings of English- and Japanese-Speaking Participants

In order to test whether English- and Japanese speaking participants differed in their emotion ratings, we applied analyses of variance (ANOVAs) to test for potential group differences. Results showed no significant differences between the ratings of English- and Japanese-speaking participants for the Expressive Condition (ratings for Joy:  $M_{\text{English}} = 3.81$ ,  $SD_{\text{English}} = 1.27$ ,  $M_{\text{Japanese}} = 3.87$ ,  $SD_{\text{Japanese}} = 1.35$ ,  $F_{(1, 270)} = 0.14$ ,  $p = 0.71$ ,  $\eta_p^2 = 0.001$ ; ratings for Sadness:  $M_{\text{English}} = 4.16$ ,  $SD_{\text{English}} = 1.33$ ,  $M_{\text{Japanese}} = 4.27$ ,  $SD_{\text{Japanese}} = 1.39$ ,  $F_{(1, 270)} = 0.46$ ,  $p = 0.50$ ,  $\eta_p^2 = 0.002$ ). The same was the case for the Neutral Condition (ratings for Joy:  $M_{\text{English}} = 3.73$ ,  $SD_{\text{English}} = 1.17$ ,  $M_{\text{Japanese}} = 3.81$ ,  $SD_{\text{Japanese}} = 1.26$ ,  $F_{(1, 262)} = 0.26$ ,  $p = 0.61$ ,  $\eta_p^2 = 0.001$ ; ratings for Sadness:  $M_{\text{English}} = 4.01$ ,  $SD_{\text{English}} = 1.15$ ,  $M_{\text{Japanese}} = 4.25$ ,  $SD_{\text{Japanese}} = 1.24$ ,  $F_{(1, 262)} = 2.61$ ,  $p = 0.11$ ,  $\eta_p^2 = 0.01$ ).

In addition, we monitored for differences between participants with and without experience in analyzing linguistic audio-material. Applying again analyses of variance, results showed no significant differences between the ratings of Japanese-speaking participants with and without experience within the Expressive Condition (ratings for Joy:  $M_{\text{experienced}} = 3.56$ ,  $SD_{\text{experienced}} = 0.84$ ,  $M_{\text{unexperienced}} = 3.89$ ,  $SD_{\text{unexperienced}} = 1.29$ ,  $F_{(1, 18)} = 0.50$ ,  $p = 0.48$ ,  $\eta_p^2 = 0.003$ , ratings for Sadness:  $M_{\text{experienced}} = 3.27$ ,  $SD_{\text{experienced}} = 0.86$ ,  $M_{\text{unexperienced}} = 3.40$ ,  $SD_{\text{unexperienced}} = 1.40$ ,  $F_{(1, 18)} = 0.07$ ,  $p = 0.80$ ,  $\eta_p^2 = 0.000$ ). The same was the case for English-speaking participants with and without analyses-experience in the Expressive Condition (ratings for Joy:  $M_{\text{experienced}} = 3.78$ ,  $SD_{\text{experienced}} = 1.27$ ,  $M_{\text{unexperienced}} = 3.82$ ,  $SD_{\text{unexperienced}} = 1.39$ ,  $F_{(1, 16)} = 0.40$ ,  $p = 0.85$ ,  $\eta_p^2 = 0.000$ ; ratings for Sadness:  $M_{\text{experienced}} = 3.91$ ,  $SD_{\text{experienced}} = 1.62$ ,  $M_{\text{unexperienced}} = 3.73$ ,  $SD_{\text{unexperienced}} = 1.32$ ,  $F_{(1, 16)} = 0.42$ ,  $p = 0.52$ ,  $\eta_p^2 = 0.003$ ). Analyses of the Neutral Condition likewise showed no significant differences between experienced and unexperienced Japanese and/or English participants (all  $p \geq 0.29$ ,  $\eta_p^2 \geq 0.010$ ). Comparing in a further step all Non-German speaking participants (i.e., English and Japanese participants) with and without experience in analyzing linguistic audio-material, results again showed no significant differences, this applied to both the Expressive and the Neutral Condition (all  $p \geq 0.15$ ,  $\eta_p^2 \geq 0.008$ ).

Based on these results, we henceforth considered English- and Japanese-speaking participants in all subsequent analyses jointly as non-German-speaking participants.

### Statistical Analyses

We performed linear mixed effect analyses using the package lme4 (Bates et al., 2014) in the statistical software R (R Core Team, 2013) to predict emotion ratings, i.e., participants’ ratings for Joy and Sadness. *P*-values were obtained using lmerTest (Kuznetsova et al., 2015). All other analyses were conducted in SPSS (IBM SPSS Statistics for Windows, Version 22.0, IBM Corp., 2013).

For all models, we included the factors *language* (non-German-speaking participants vs. German-speaking participants) and *poem preclassification* (i.e., the poems' preclassification as either joyful or sad), together with the random intercept factors *participant* and *poem*. In addition, the Expressive Condition allowed us to include the within-participant factor *prosody* (joy-supporting vs. sadness-supporting). In a second step, we replaced this additional factor by *match/mismatch* (match vs. mismatch between prosodic features and semantic content e.g., poems that were preclassified as joyful were presented with a sadness-supporting prosody). The initial models were always specified as full factorial, i.e., as including all possible interaction terms. Subsequent models focused on specific effects when qualified by significant interactions.

## Expressive Condition

### Full model

The models that were used to predict Joy ratings showed main effects of *poem preclassification* ( $b = 3.53$ ,  $SE = 0.27$ ,  $t = 13.20$ ,  $p \leq 0.001$ ) and of *language* ( $b = 1.35$ ,  $SE = 0.15$ ,  $t = 8.92$ ,  $p \leq 0.001$ ). Further, analyses revealed an interaction between *poem preclassification* and *language* ( $b = -2.77$ ,  $SE = 0.21$ ,  $t = -13.27$ ,  $p \leq 0.001$ ). Importantly, there was no main effect of *prosody* ( $b = 0.12$ ,  $SE = 0.10$ ,  $t = 1.11$ ,  $p = 0.27$ ), but a significant interaction of *prosody* and *language* ( $b = 0.41$ ,  $SE = 0.21$ ,  $t = 1.98$ ,  $p = 0.05$ ). The prediction of participants' Sadness ratings showed the same results. Again, we found main effects for *poem preclassification* ( $b = -3.07$ ,  $SE = 0.28$ ,  $t = -11.03$ ,  $p \leq 0.001$ ) and *language* ( $b = -1.14$ ,  $SE = 0.18$ ,  $t = -6.23$ ,  $p \leq 0.001$ ), as well as an interaction of *poem preclassification* and *language* ( $b = 2.10$ ,  $SE = 0.24$ ,  $t = 8.84$ ,  $p \leq 0.001$ ), and of *prosody* and *language* ( $b = -0.52$ ,  $SE = 0.24$ ,  $t = -2.20$ ,  $p = 0.03$ ). Again, there was no main effect of *prosody* ( $b = -0.15$ ,  $SE = 0.12$ ,  $t = -1.27$ ,  $p = 0.20$ ).

Replacing the factor *prosody* by *match/mismatch*, Joy rating showed main effects of *poem preclassification* ( $b = 3.41$ ,  $SE = 0.27$ ,  $t = 12.76$ ,  $p \leq 0.001$ ) and of *language* ( $b = 1.76$ ,  $SE = 0.15$ ,  $t = 11.68$ ,  $p \leq 0.001$ ). We found an interaction between *poem preclassification* and *language* ( $b = -3.19$ ,  $SE = 0.21$ ,  $t = -15.26$ ,  $p \leq 0.001$ ). *Match/mismatch* showed no main effect ( $b = -0.12$ ,  $SE = 0.11$ ,  $t = -1.11$ ,  $p = 0.27$ ), but a significant interaction with *language* ( $b = -0.41$ ,  $SE = 0.21$ ,  $t = 1.98$ ,  $p = 0.05$ ).

The model predicting Sadness ratings revealed a main effect for *poem preclassification* ( $b = -2.92$ ,  $SE = 0.28$ ,  $t = -10.48$ ,  $p \leq 0.001$ ) and *language* ( $b = -1.66$ ,  $SE = 0.18$ ,  $t = -9.09$ ,  $p \leq 0.001$ ), as well as an interaction of *poem preclassification* and *language* ( $b = 2.62$ ,  $SE = 0.24$ ,  $t = 11.05$ ,  $p \leq 0.001$ ). The factor *match/mismatch* showed no main effect ( $b = 0.52$ ,  $SE = 0.12$ ,  $t = 1.27$ ,  $p = 0.20$ ), but a significant interaction with the factor *language* ( $b = 0.52$ ,  $SE = 0.24$ ,  $t = -2.20$ ,  $p = 0.03$ ; for decompositions of these interactions, see below).

### Non-German-speaking participants

The decomposition of the interactions per language-group (Non-German-speaking vs. German-speaking participants)

revealed that both emotion ratings of the non-German-speaking participants were significantly influenced by the factor *prosody* (ratings for Joy:  $b = 0.46$ ,  $SE = 0.22$ ,  $t = 2.11$ ,  $p = 0.04$ ; ratings for Sadness:  $b = -0.63$ ,  $SE = 0.22$ ,  $t = -2.83$ ,  $p = 0.01$ ). Also, analyses showed an effect of *poem preclassification* (ratings for Joy:  $b = 0.68$ ,  $SE = 0.22$ ,  $t = 3.11$ ,  $p \leq 0.01$ ; ratings for Sadness:  $b = -0.91$ ,  $SE = 0.26$ ,  $t = -3.44$ ,  $p = 0.01$ ).

When replacing *prosody* by *match/mismatch*, analyses showed that the emotion rating of the non-German participants showed a significant effect of *match/mismatch* (ratings for Joy:  $b = -0.46$ ,  $SE = 0.22$ ,  $t = -2.11$ ,  $p = 0.04$ ; ratings for Sadness:  $b = 0.63$ ,  $SE = 0.22$ ,  $t = 2.83$ ,  $p = 0.01$ ), but not of *poems preclassification* (ratings for Joy:  $b = 0.22$ ,  $SE = 0.22$ ,  $t = 1.00$ ,  $p = 0.32$ ; ratings for Sadness:  $b = -0.27$ ,  $SE = 0.26$ ,  $t = -1.04$ ,  $p = 0.06$ ).

### German-speaking control group

In contrast to the results for the non-German speaking participants, native German-speaking listeners' ratings were only significantly influenced by the factor *poem preclassification* (ratings for Joy:  $b = 3.52$ ,  $SE = 0.34$ ,  $t = 10.71$ ,  $p \leq 0.001$ ; ratings for Sadness:  $b = -3.06$ ,  $SE = 0.32$ ,  $t = -9.56$ ,  $p \leq 0.001$ ). They rated joyful poems as more joyful ( $N = 400$ ,  $M = 5.49$ ,  $SD = 1.09$ ) than sad poems ( $N = 400$ ,  $M = 2.05$ ,  $SD = 0.98$ ,  $F_{(1, 798)} = 2196.22$ ,  $p \leq 0.001$ ,  $\eta^2 = 0.73$ ) and sad poems as sadder than joyful ones ( $M = 5.27$ ,  $SD = 1.34$ ,  $F_{(1, 798)} = 1,821.48$ ,  $p \leq 0.001$ ,  $\eta^2 = 0.59$ ). The presentation of joyful and sad poems with either a joy- or a sadness-supporting prosody did not significantly affect the emotion ratings of the German-speaking listeners (ratings for Joy:  $b = -0.17$ ,  $SE = 0.13$ ,  $t = -1.28$ ,  $p = 0.20$ ; ratings for Sadness:  $b = 0.09$ ,  $SE = 0.16$ ,  $t = 0.57$ ,  $p = 0.57$ ).

Analyses of the models that included the factor *match/mismatch* instead of *prosody* revealed that German-speaking participants' emotion ratings were not influenced by the factor *match/mismatch* (ratings for Joy:  $b = -0.11$ ,  $SE = 0.09$ ,  $t = -1.18$ ,  $p = 0.20$ ; ratings for Sadness:  $b = 0.15$ ,  $SE = 0.11$ ,  $t = 1.31$ ,  $p = 0.20$ ). Rather, their ratings showed, again, an effect of *poem preclassification* (ratings for Joy:  $b = 3.41$ ,  $SE = 0.33$ ,  $t = 10.37$ ,  $p \leq 0.001$ ; ratings for Sadness:  $b = -2.91$ ,  $SE = 0.32$ ,  $t = -9.05$ ,  $p \leq 0.001$ ).

## Neutral Condition

### Full model

As in the Expressive Condition, the results for the Neutral Condition showed a main effect of *poem preclassification* (ratings for Joy:  $b = 3.36$ ,  $SE = 0.28$ ,  $t = 11.90$ ,  $p \leq 0.001$ ; ratings for Sadness:  $b = -3.03$ ,  $SE = 0.24$ ,  $t = -12.41$ ,  $p \leq 0.001$ ) as well as *language* (ratings for Joy:  $b = 1.28$ ,  $SE = 0.11$ ,  $t = 11.26$ ,  $p \leq 0.001$ ; ratings for Sadness:  $b = -0.83$ ,  $SE = 0.14$ ,  $t = -5.83$ ,  $p \leq 0.001$ ) on participants' emotion ratings. Further, our analyses again revealed interactions between these two independent variables (ratings for Joy:  $b = -2.64$ ,  $SE = 0.14$ ,  $t = -18.38$ ,  $p \leq 0.001$ ; ratings for Sadness:  $b = 2.28$ ,  $SE = 0.16$ ,  $t = 13.94$ ,  $p \leq 0.001$ ).

The decomposition of the interaction between *poem preclassification* and *language* showed that the non-German-speaking participants clearly differed in their ratings from

the German-speaking group. A separate inspection of the mean values of emotion rating for the joyful and sad poems showed that non-German-speaking participants rated joyful poems as less joyful ( $M = 4.15, SD = 1.20$ ) and sadder ( $M = 3.54, SD = 1.29$ ) than the German-speaking group ( $M_{Joy} = 5.48, SD_{Joy} = 1.16, M_{Sadness} = 2.09, SD_{Sadness} = 1.13$ , all  $F_{(1, 549)} \geq 134.42$ , all  $p \leq 0.001$ , all  $\eta^2 \geq 0.20$ ). Likewise, the non-German-speaking participants rated sad poems as more joyful ( $M = 3.41, SD = 1.14$ ) and less sad ( $M = 4.28, SD = 1.20$ ) than the German-speaking participants ( $M_{Joy} = 2.12, SD_{Joy} = 0.99, M_{Sadness} = 5.12, SD_{Sadness} = 1.40$ , all  $F_{(1, 550)} \geq 37.95$ , all  $p \leq 0.001$ , all  $\eta^2 \geq 0.07$ ; see **Figure 2** for an illustration using the mean values for emotions ratings of German and non-German speaking participants for poems preclassified as joyful and sad).

**Non-German-speaking participants**

The factor of *poem preclassification* again showed an influence on the ratings of the non-German-speaking listeners (ratings for Joy:  $b = 0.72, SE = 0.17, t = 4.10, p = 0.01$ ; ratings for Sadness:  $b = -0.74, SE = 0.21, t = -3.47, p = 0.05$ ).

**German-speaking control group**

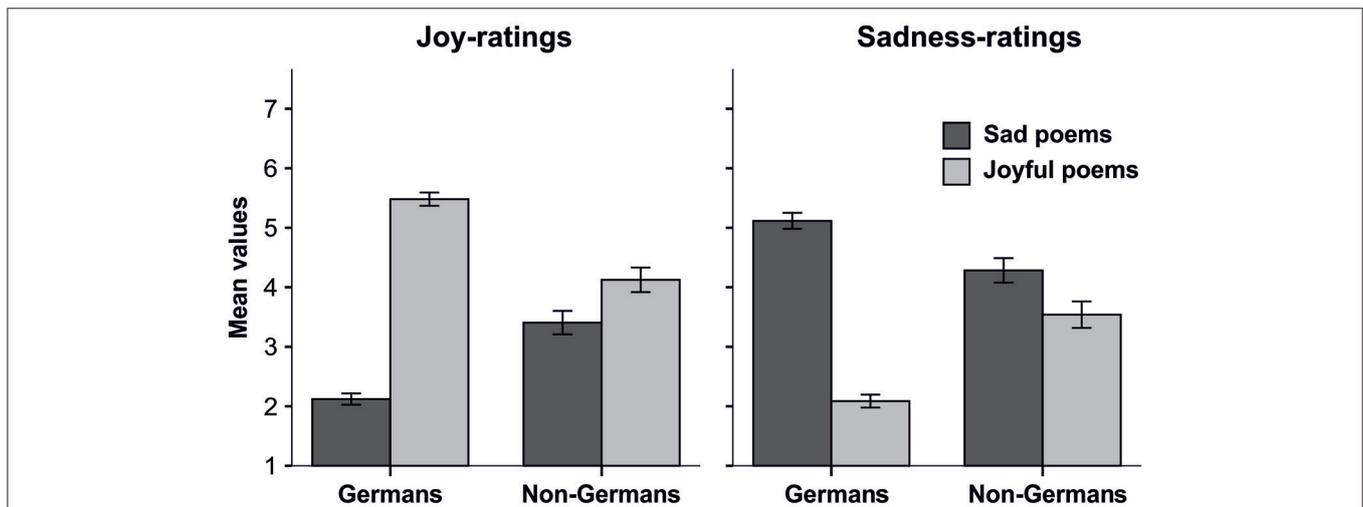
The effect of *poem preclassification* held for the German-speaking listeners in the Neutral Condition. Therefore, joyful poems were rated as more joyful, and sad poems were rated as sadder (ratings for Joy:  $b = 3.36, SE = 0.35, t = 9.66, p \leq 0.001$ ; ratings for Sadness:  $b = -3.03, SE = 0.29, t = -10.58, p \leq 0.001$ ).

**DISCUSSION**

In line with the hypothesis of prosody-driven sound-emotion associations in poetry, we found a significant relation between prosodic features of participants' recitations of poems and their ratings of joy and sadness: On average, pitch and tempo

values were higher for joyful than for sad poems across German native speakers (Experiment 1). Our cross-language perception experiment (Experiment 2) further illustrated that pitch and tempo values are indeed interpreted with respect to emotional content: non-German-speaking participants without semantic access to the poems were particularly sensitive to the prosodic manipulation, i.e., they rated poems with a joy-supporting prosody as more joyful and poems with a sadness-supporting prosody as sadder. Further, our analyses confirmed the expectation that, in the Expressive Condition, non-German-speaking participants were sensitive to the difference between poem versions in which emotional content and prosodic features matched, and poems in which emotional content and prosodic features did not match. At the same time, however, German participants did not show significantly different ratings dependent on our prosodic manipulations. Rather, the content-based preclassification of the poems as either joyful or sad was the only significant predictor of their ratings, independently of whether prosodic features were in line with the semantic content of the poems or not. Thus, in addition to the results of our experiments, our investigation also highlights the importance of semantic access to the content of a poem, which, if available, has to be considered the most vital predictor of emotion perception in poetry.

In the neutral condition, as expected, non-German-speaking participants' and German-speaking-participants' ratings differed significantly. However, and rather unexpectedly, the content-based preclassification of the poems as joyful or sad also predicted the emotion ratings of the non-German-speaking listeners in the neutral condition, albeit to a weaker degree than for the German control group. Given the result from the Expressive Condition that non-German-speaking listeners were sensitive to the difference between match and mismatch of prosodic features and semantic content, the correlation between preclassification and emotion ratings of the non-German group is likely not to be based on linguistic comprehension. Rather, the likelihood of



**FIGURE 2** | Bar charts showing mean values of Joy and Sadness- ratings from German and non-German speaking participants, separately displayed for joyful and sad poems. Error bars indicate the 95% confidence intervals.

non-German-speaking participants to detect the correct emotion may rather derive from the fact that the presented poems retained prosody-based characteristic of joy and sadness beyond the parameters investigated here. As stated in the Stimulus section, the presented poems retained their original pitch contours and we exclusively defined and modified emotional prosody in terms of mean pitch and mean articulation rate. Clearly, these measures hide a large amount of important variance, and they cannot account at all for the dynamic contours of emotional prosody throughout intonation phrases, which are likely to make a strong contribution to the overall perception of emotional prosody (e.g., Pierrehumbert, 1980; Gussenhoven, 2004; Ladd, 2008).

The non-inclusion of the pitch contour characteristics in the experimental modification constitutes a general limitation of our study. Future studies are thus called for that not only consider pitch contour and additional cues of emotional prosody, but also investigate suprasegmental features separately and in a step-wise addition. Also, further evidence is needed to examine prosodic effects and their influence on emotion perception during poetry reading by native recipients and to investigate to what extent content-based emotion perception also relies on prosodic feature during (silent) reading. Furthermore, future studies should include stimuli that cannot as clearly be assigned to dominant emotions as it was the case for the poems used in our experiments. Additional variables could include poetic forms such as sound poems, or mixed emotions such as nostalgia or melancholy.

Notwithstanding these limitations, our study offers a prosody-based account for sound-emotion associations in poetry that goes beyond the repeatedly and inconsistently tested assumption of a relation between phoneme frequencies and emotion perception. Our results strongly indicate that emotional prosody plays a substantial role in sound-emotion associations in poetry, if content-based semantic access is unavailable. The

suprasegmental features investigated here are an important factor for recitations by native speakers, and also influence the emotion ratings of listeners who have no access to the content of the presented poems. Thus, emotional prosody is an important dimension in both expressing and deciphering the emotional content of poems. This is all the more important since purely semantic understanding may be not as straightforward in poetry as in other textual forms, be that for reasons of the greater importance of stylistic figures, deviations from common language use, the particularly condensed forms of evoking the content, or more associative imagery.

## AUTHOR CONTRIBUTIONS

MK designed the study and compiled the poetic corpus. MK and AR decided upon experimental conditions, manipulated the audio files and gathered data. MK conducted all analyses. MK, WM, and MS interpreted the data and wrote the paper.

## FUNDING

This paper's data acquisition, analyses, and writing were conducted at and funded by the Max Planck Institute for Empirical Aesthetics in Frankfurt am Main, Germany.

## ACKNOWLEDGMENTS

We would like to thank Freya Materne, R. Muralikrishnan, and Dominik Thiele for their support in data acquisition and preparation, Yasuhiro Sakamoto for translating the online questionnaire into Japanese, Erik Stahnke for the recordings of our acoustic stimulus material, Fabian Zweck for the loudness normalization of the material, and Cornelius Abel for his help with the figures.

## REFERENCES

- Aitken, R. C. (1969). Measurement of feelings using visual analogue scales. *Proc. R. Soc. Med.* 62, 989–993.
- Albers, S. (2008). *Lautsymbolik in Ägyptischen Texten*. Mainz: Philipp Von Zabern Verlag.
- Alter, K. (2002). "Suprasegmentale Merkmale und Prosodie," in *Arbeitsbuch Linguistik*, ed H. M. Müller (Paderborn: UTB-Ferdinand Schöninghausen), 148–169.
- Aryani, A., Kraxenberger, M., Ulrich, S., Conrad, M., and Jacobs, A. M. (2016). Measuring the basic affective tone of poems via phonological saliency and iconicity. *Psychol. Aesthetics Creativ. Arts* 10, 191–204. doi: 10.1037/aca0000033
- Auracher, J., Albers, S., Zhai, Y., Gareeva, G., and Stavniychuk, T. (2010). P is for happiness, N is for sadness: universals in sound iconicity to detect emotions in poetry. *Discourse Process* 48, 1–25. doi: 10.1080/01638531003674894
- Banse, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70:614. doi: 10.1037/0022-3514.70.3.614
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear Mixed-Effects Models Using Eigen and s4*. R package version, 1.
- Belkner (2014). *R128GAIN*. Version 3. Available online at: <http://r128gain.sourceforge.net/>
- Boersma, P., and Weenink, D. (2015). *Praat: Doing Phonetics by computer [Computer Program]*. Version 6.0.
- Demmerling, C., and Landweer, H. (2007). *Philosophie der Gefühle: Von Achtung bis Zorn*. Stuttgart: Metzler.
- Ekman, P. (1992). An argument for basic emotions. *Cogn. Emotion* 6, 169–200.
- Ekman, P., and Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion Rev.* 3, 364–370. doi: 10.1080/02699939208411068
- Fónagy, I. (1961). Communication in Poetry. *Word* 17, 194–218. doi: 10.1080/00437956.1961.11659754
- Funke, F., and Reips, U. D. (2012). Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field Methods* 24, 310–327. doi: 10.1177/1525822X12444061
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Hegel, G. (1986). *Vorlesungen über Ästhetik*. Frankfurt am Main: Suhrkamp.
- Hirst, D. J. (2011). The analysis by synthesis of speech melody: from data to models. *J. Speech Sci.* 1, 55–83.
- International Telecommunication Union (2012). *ITU-R BS.1770-3: Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level*. (Geneva)
- Jack, R. E., Garrod, O. G., and Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Curr. Biol.* 24, 187–192. doi: 10.1016/j.cub.2013.11.064
- Jakobson, R. (1960). Closing statement: linguistics and poetics. *Style Lang.* 350:377.
- Jakobson, R., and Waugh, L. R. (1979/2002). *The Sound Shape of Language*. Berlin: Walter de Gruyter.

- Jessen, M. (2007). Forensic reference data on articulation rate in German. *Sci. Justice* 47, 50–67. doi: 10.1016/j.scijus.2007.03.003
- Kaiser, L. (1962). Communication of affects by single vowels. *Synthese* 14, 300–319. doi: 10.1007/BF00869311
- Kraxenberger, M., and Menninghaus, W. (2016a). Mimological Reveries? Disconfirming the Hypothesis of Phono-Emotional Iconicity in Poetry. *Front. Psychol.* 7:1779. doi: 10.3389/fpsyg.2016.01779
- Kraxenberger, M., and Menninghaus, W. (2016b). Emotional Effects of Poetic Phonology, Word Positioning and Dominant Stress Peaks in Poetry Reading. *Sci. Stud. Lit.* 6, 298–313. doi: 10.1075/ssol.6.1.06kra
- Künzel, H. J. (1997). Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguist.* 4, 48–83.
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). *Package 'lmerTest'*. R package version, 2.0.
- Ladd, D. R. (2008). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Lüdtke, J., Meyer-Sickendieck, B., and Jacobs, A. M. (2014). Immersing in the stillness of an early morning: testing the mood empathy hypothesis of poetry reception. *Psychol. Aesthet. Creativ. Arts* 8, 363–377. doi: 10.1037/a0036826
- Menninghaus, W., Wagner, V., Wassiliwizky, E., Jacobsen, T., and Knoop, C. A. (2017). The emotional and aesthetic powers of parallelistic diction. *Poetics* 63, 47–59. doi: 10.1016/j.poetic.2016.12.001
- Paulmann, S. (2006). *Electrophysiological Evidence on the Processing of Emotional Prosody: Insights from Healthy and Patient Populations*. Doctoral dissertation, Max Planck Institute for Human Cognitive and Brain Sciences Leipzig.
- Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., and Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: a comparison of four languages. *J. Phon.* 37, 417–435. doi: 10.1016/j.wocn.2009.07.005
- Pfützinger, H. R. (2001). *Phonetische Analyse der Sprechgeschwindigkeit. Forschungsberichte des Instituts für Phonetik und sprachliche Kommunikation der Universität München* 38, 117–264.
- Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. Doctoral Dissertation, Massachusetts Institute of Technology.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*.
- Reichel, U. D. (2012). “PermA and balloon: tools for string alignment and text processing,” in *Proceedings of Interspeech*, Vol. 346 (Portland, OR).
- Reichel, U. D., and Kisler, T. (2014). “Language-Independent grapheme-phoneme conversion and word stress assignment as a web service,” in *Elektronische Sprachverarbeitung: Studententexte zur Sprachkommunikation*, Vol. 71, ed R. Hoffmann (Dresden: TUDpress), 42–49.
- Russell, J. A. (1980). A circumplex model of affect. *J. Person. Soc. Psychol.* 39, 1161–1178.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 110, 145–172. doi: 10.1037/0033-295X.110.1.145
- Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychol. Bull.* 99:143. doi: 10.1037/0033-2909.99.2.143
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Soc. Sci. Inf. Sur Les Sci. Sociales* 44, 695–729. doi: 10.1177/0539018405058216
- Scherer, K. R. (2013). “Emotion in action, interaction, music, and speech,” in *Language, Music, and the Brain: A Mysterious Relationship*, ed A. A. Micheal (Cambridge: MIT press), 107–139.
- Scherer, K. R., Banse, R., Wallbott, H. G., and Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motiv. Emot.* 15, 123–148. doi: 10.1007/BF00995674
- Schmitz, H. (1969). *System der Philosophie: der gefühlsraum*. Bonn: Bouvier.
- Schrott, R., and Jacobs, A. M. (2011). *Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren*. München: Hanser.
- Shklovsky, V. (1919/2012). “Art as technique,” in *Russian Formalist Criticism: Four Essays*, ed L. Lemon (Lincoln, NE: University of Nebraska Press), 3–24.
- Stolarski, L. (2015). Pitch patterns in vocal expression of “Happiness” and “Sadness” in the reading aloud of prose on the basis of selected audiobooks. *Res. Lang.* 13, 141–162. doi: 10.1515/rela-2015-0016
- Titze, I. R. (2000). *Principles of Voice Production*. Iowa City, IA: National Center for Voice and Speech.
- Trouvain, J., Koreman, J., Erriquer, A., and Zinn, B. (2001). *Articulation rate measures and their relation to phone classification in spontaneous and read German speech, Adaption-2001*, 155–158.
- Tsur, R. (1992). *What Makes Sound Patterns Expressive? The Poetic Mode of Speech Perception*. Durham, NC: Duke University Press.
- Unipark and QuestBack Globalpark AG. (2014). *Enterprise Feedback Suite*. EFS Survey. Available online at: <https://www.unipark.com/>
- Valéry, P. (1958). *Remarks on Poetry. The Art of Poetry* (Transl. D. Folliot), New York, NY: Pantheon Press.
- Van Bezooijen, R. (1984). *The Characteristics and Recognizability of Vocal Expressions of Emotion*. Dordrecht: Foris Publication.
- Ververidis, D., and Kotropoulos, C. (2006). Emotional speech recognition: resources, features, and methods. *Speech Commun.* 48, 1162–1181. doi: 10.1016/j.specom.2006.04.003
- Whissell, C. (2002). Emotion conveyed by sound in the poetry of Alfred, Lord Tennyson. *Empir. Stud. Arts* 20, 137–155. doi: 10.2190/6K4G-LWPQ-RAY8-67QG
- Whissell, C. (2011). “To those who feel rather than to those who think:” Sound and emotion in Poe’s poetry. *Int. J. Engl. Lit.* 2, 149–156.
- Winko, S. (2003). *Kodierte Gefühle: Zu einer Poetik der Emotionen in Lyrischen und Poetologischen Texten um 1900*. Berlin: Erich Schmidt Verlag.

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

Copyright © 2018 Kraxenberger, Menninghaus, Roth and Scharinger. 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.