Studying Emotive Effects in Poetry by Quantifying Open-Ended Impressions

*** Author accepted manuscript ***

Full reference:
Studying Emotive Effects in Poetry by Quantifying Open-Ended Impressions

Chen Gafni¹ and Reuven Tsur²

¹ University of Haifa, Haifa, Israel, ² Tel-Aviv University, Tel-Aviv, Israel

Corresponding author:
Chen Gafni, University of Haifa.
Email: chen.gafni@gmail.com

Abstract
Poems, as aesthetic objects, generate a subjective experience, which can be different for different readers. In this paper, we propose a method to quantify these subjective experiences. We gave participants three parallel excerpts and asked them to describe, in free text, the perceived emotive qualities of these excerpts. The descriptions were analysed quantitatively according to the dimensions of the Valence-Arousal-Dominance model of emotion. With the help of additional rating tasks and a structural theory of phonetic symbolism, we attribute the perceived emotive qualities to an interaction between the meaning of words, patterns of alliteration, and metric deviation.

Keywords
Sound symbolism, sound iconicity, cognitive poetics, aesthetics, dimensional models of emotion, reader’s response
A poem has a verbal structure, a meaning and a perceived quality. Meaning and poetic structure can be described in quite objective terms. But perceived quality crucially depends on reader response. A perceived quality is a regional quality emerging from the interaction of a wide range of semantic, phonetic, syntactic and thematic variables. Various readers realise different subgroups of these variables and, consequently, report a wide range of perceived effects. In a quest for “scientific” results, most researchers isolate a few variables from the complex, and force their participants to respond to them, hoping to discover the perceived quality of the text. For example, a common practice in the field is to have na"ive readers rate poems relative to some predefined terms such as joy, sad, friendly (Aryani, Kraxenberger, Ullrich, Jacobs, & Conrad, 2016; Auracher, Albers, Zhai, Gareeva, & Stavniychuk, 2011; Kraxenberger & Menninghaus, 2016). This is often followed by examining the correlations between the subjective ratings and some global properties of the poem, such as the frequency distribution of certain speech sounds (e.g. examining whether the “sadness” of poems correlates with the frequency of nasal consonants in the texts).

The methodology described above is popular in studies focusing on the relation between the sound content of the poem and its perceived quality. However, while this methodology is appropriate for testing specific hypotheses it also has certain disadvantages. First, the assumed relation between speech sounds and emotions is often based solely on statistical correlations with no proper theory to account for these correlations. This limits the ability to generalise and compare the results across studies. Second, the use of predefined emotive terms confines the investigation to the chosen dimensions and leads to an incomplete evaluation of the aesthetic qualities of the poems. This could be the case if the terms provided by the researchers are different from the ones participants would use if they were asked to describe the perceived effect in their own words (see Scherer, 2005).

The goal of the current study was to establish a more suitable methodology for studying the emotive quality of poems and the contribution of speech sounds to the perceived quality. The proposed methodology relies mainly on free verbal responses that provide the complete and unguided subjective impressions of the readers. In addition to a qualitative analysis of the responses, this paper proposes a method to

1 The use of global measures to characterise whole poems is also problematic but the issue is beyond the scope of this paper.
quantify the emotive dimensions of the responses using a dictionary of affect. The analysis of verbal responses can be complemented by results of additional tasks, such as statement rating, designed to test specific hypotheses. We present our theory and practice through an investigation of phonetic symbolism. Finally, an account for the contribution of sound patterns to the perceived quality of the text should be based on careful analysis guided by linguistic theory of phonetic symbolism. In the following sections, we describe the theoretical frameworks we used in this study.

Analysing the emotive quality of poems
One of the main goals of the current study was to develop a method for obtaining the emotive quality of poetic text that is: (i) quantifiable, (ii) generalisable across texts, and (iii) based on the unguided, free impressions of readers. The latter criterion is extremely important. In previous studies, participants were asked to evaluate the emotive quality of poems along certain pre-determined dimensions. However, such methodology fails to capture the full aesthetic experience, as it forces participants to focus on those dimensions the researchers were interested in.

To remedy this “researcher bias”, we claim that emotive evaluations of poetic texts (or any other aesthetic objects, for that matter) should be based primarily on open-ended responses. Such responses provide more details on how readers truly perceive the object of interest. However, the verbal responses need to be transformed into quantitative measures to allow for comparisons of various texts. This can be done using dimensional models of emotion.

Dimensional models of emotion represent words with emotive content in terms of some basic dimensions. In the current study, we used the Valence-Arousal-Dominance (VAD) model2 (Bakker, van der Voordt, Vink, & de Boon, 2014; Mehrabian & Russell, 1974; Russell & Mehrabian, 1977). Words are located in this three dimensional space according to their values along each of the three bipolar scales: valence (positive–negative/pleasant–unpleasant emotions), arousal (active–passive/high-aroused–low-aroused emotions), and dominance (dominant–submitive/controlling–controlled). The exact coordinates of each word are determined by judgments collected from a large

---

2 Different authors used somewhat different terminologies for their basic axes (e.g. pleasure instead of valence), which correspond completely or partially to the VAD dimensions (see review in Bakker et al., 2014).
number of participants, who are asked to evaluate the words along several scales (the scales could be the minimal set of the basic dimensions of the model or a larger set of scales assumed to represent the basic dimensions). The outcome of this procedure is an affective lexicon, in which words are assigned numerical values representing their emotive quality. These values can be utilised to obtain a quantitative measure of the emotive quality of aesthetic objects based on content words extracted from open-ended responses to the object.

A structural model of phonetic symbolism

The current paper demonstrates a method for studying the emotive effects of phonetic symbolism (or sound symbolism) in poetry. But, before discussing the theoretical model of phonetic symbolism, we should explain what we mean by “phonetic symbolism” and how it is related to the similar term of sound iconicity. According to Trask (1996), sound symbolism is “Any of various types of iconic (non-arbitrary) relations between the sound of a word and its meaning”. Thus, according to this definition, symbolism and iconicity are synonymous, both meaning non-arbitrary. Indeed, some scholars seem to use these terms “quasi-synonymously” (see Auracher et al., 2011: 2). However, there is an alternative view according to which, symbolism is not synonymous with iconicity. According to Hoshi, Kwon, Akita, & Auracher (2019: 2): “in the Peircean tradition of semiotics, the term symbolism is reserved for signs whose relation to their meaning is arbitrarily imputed (Peirce, 1867, p. 294); we prefer the term sound iconicity – that is, a relation between sound and meaning based on intrinsic resemblance (Wescott, 1971).”

We conceive of phonetic symbolism as a non-fixed relation between sound and meaning that is generated by an unpredictable (but not random) interaction of phonetic and semantic features. Importantly, these features constrain rather than determine the result. Therefore, given the above distinction between “symbolism” and “iconicity”, we prefer to use the term “symbolism” (see also Tsur, 2019). Nevertheless, our approach is compatible with the view of iconicity as semiotic relations, which builds on sign-object structural analogies that can be mediated by context (Dingemanse, Perlman, & Perniss, 2020). Thus, it is not important whether one calls it “sound symbolism” or
“sound iconicity”, as long as it is understood as a structural relationship between sign and object and not as an inherent property of the sign.³

To demonstrate the difference between the structural approach and the inherent property approach to sound symbolism/iconicity, consider the study by Auracher et al. (2011). In this cross-cultural, cross-linguistic study, participants were asked to read poems and rate the appropriateness of given emotion words (e.g. sadness, happiness) referring to the feelings expressed in the poem. It was found that the appropriateness ratings of emotion words were significantly correlated with the ratio of plosive versus nasal consonants in the texts. Auracher et al. (2011: 21) conclude that “there is a universal tendency to express happy and active feelings with plosive sounds, whereas sad and passive feelings are encoded in nasal sounds.” Thus, Auracher et al. seem to support the notion of inherent iconicity where speech sounds have intrinsic emotive qualities. We, by contrast, argue for a more flexible and complex view of sound-meaning interaction. More importantly, we are not satisfied with merely establishing statistical correlations between the frequency of certain speech sounds and emotive qualities. Rather, we attempt to discover what makes certain sounds more appropriate for expressing certain emotions.

In order to understand the contribution of sound patterns to the emotive quality of a poem, one needs a structural theory that projects emotions and speech sounds onto common dimensions. Tsur & Gafni (2019a, 2019b) propose such a theory, whose main principles are: (i) speech sounds, meanings and emotions can be represented as bundles of features; (ii) effects of phonetic symbolism occur when structural similarity is detected between sound and meaning; (iii) speech sounds can express only elementary perceptions, not complex emotions or meanings; (iv) complex effects are generated by interacting with meaning: sound and meaning may transfer features to each other, activate them in each other, or reject them by some incompatible feature; (vi) such interactions are not a pre-defined, inherent property, but rather the result of creative cognitive processes that form unpredictable connections between sound and meaning features, after the event.

³ This is what the Roman rhetorician Quintilianus (c. 35–100) has to say on this issue: “For it makes no difference by which name is either called, so long as its stylistic value is apparent, since the meaning of things is not altered by a change of name. For just as men remain the same, even though they adopt a new name, so these artifices will produce exactly the same effect, whether they are styled tropes or figures, since their values lie not in their names, but in their effect” (Quintilian, 1920: 353).
An important consequence of the structural theory is that speech sounds do not have inherent valence. The same speech sounds can support, in different contexts, both positive and negative emotions. In fact, of the three basic dimensions of emotion, the dimension of arousal is the only one directly involved in sound-emotion interactions (Aryani, Conrad, Schmidtke, & Jacobs, 2018; Aryani, Hsu, & Jacobs, 2018; Aryani et al., 2016; Bachorowski, 1999; Bänziger & Scherer, 2005; Sauter, Eisner, Calder, & Scott, 2010; Ullrich, Aryani, Kraxenberger, Jacobs, & Conrad, 2017; Ullrich, Kotz, Schmidtke, Aryani, & Conrad, 2016). High-aroused emotional states are characterised by more erratic behaviour and fast reactions to the environment. Likewise, plosive consonants (e.g. [k], [p], [t]) are characterised by abrupt changes in the acoustic waveform, and have an explosive and hard potential. Thus, plosives are natural candidates to support high-aroused states such as aggression (negative emotions) but also joy (positive emotions). On the other hand, continuous, periodic consonants (e.g. [l], [m], [n]) are characterised by a smoother waveform. Thus, they are more suitable to represent low-aroused, tender emotions such as love and grief.

As we said, speech sounds can be regarded as bundles of features. Different contexts can foreground different features of the same speech sounds for different perceptual and emotive effects. This is called double-edgedness. For example, the sibilants [s] and [ʃ] can be described both as irregular and continuous. These characters can be exploited for various purposes from imitating noises of friction (e.g. rustling leaves) to generating a calm atmosphere, like a mother hushing her crying baby.

In addition, it is important to note that, in most speech acts, speech sounds are not used iconically but, rather, as arbitrary signifiers of meaning (e.g. the sounds of the word dog are not iconic of the meaning of the concept DOG). Treating speech sounds as arbitrary signifiers of meaning is characteristic of the speech mode of auditory perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). In this mode, sound information that transmits speech is recoded by the listener into phonetic categories (or phonemes), such as [b], [k], and [m]. Crucially, the acoustic information underlying the formation of the phonetic categories is excluded from awareness in the speech mode. By contrast, in the poetic mode of speech perception (Tsur, 1992), some

---

4 Despite the use of the term “double”, speech sounds can support multiple emotional states.
5 Sibilants (also known as stridents) are a sub-class of fricative and affricate consonants produced with a concentration of energy at high frequencies. The class includes consonants such as [s], [ʃ] (as the sequence “sh” in shine), [z], and [ç] (as the sequence “ch” in chill).
of the rich precategorical auditory information is activated in awareness by the poetic context. When listeners operate in this mode they are able to attend to the perceptual properties of speech sounds beyond their role as arbitrary signifiers of meaning. Thus, it is not the phonemic categories themselves but, rather, the precategorical auditory properties underlying these categories that are responsible for effects of phonetic symbolism. Yet, the amount of available precategorical auditory information depends on the category. Plosives are highly-encoded speech sounds, in which little precategorical auditory information is available, while continuous sounds (fricatives, nasals, vowels) are lowly-encoded and, thus, allow more precategorical auditory information to reach awareness.

Finally, it is important to distinguish between lexicalised and non-lexicalised instances of phonetic symbolism. Some words contain a speech sound (or several sounds) that has some feature that is iconic of one salient feature of the meaning. These are cases of lexicalised phonetic symbolism. For example, the vowel [i] is used cross-linguistically to express the concept of smallness (e.g. *this* and *here* express a small distance from the speaker as opposed to their counterparts *that* and *there*). On the other hand, poetic texts often display ad hoc effects of phonetic symbolism that arise from a complex interaction of the meaning of the words and the sound content of the text. For example, words with a harsh, high-aroused meaning can activate the abrupt potential of plosives in other words in the text, which in turn, amplify the harsh atmosphere of the text. By contrast, low-aroused meanings are less compatible with the perceptual effect of plosives. If one component (meaning or sound) is more salient in the context, it can override the effect of the other. But a context loaded with both low-aroused words and plosives can evoke a sense of conflict. In the current study, we investigate how the emotive quality of poetic text is affected when the perceptual potential of the sound content is more compatible or less compatible with the semantic component.

**Our study**

Our starting point is Benjamin Hrushovski’s (1980) article on the relationships of sound and meaning in poetry. He argues among other things that sibilants may both imitate noises and express some hushed mood in poetry. Elsewhere (Tsur, 1992; Tsur & Gafni, 2019b) we demonstrate through acoustic feature analysis how the same sibilants can be associated with such disparate effects. The present discussion is focussed on one of
Hrushovski’s examples, two lines from Shakespeare’s sonnet 30, in which the sibilants are said to express a silent mood:

1. When to the sessions of sweet silent thought
   I summon up remembrance of things past,

Hrushovski says that these lines realise the hushing potential of the sibilants, which is focused on “sweet silent”. The use of sibilants to support a psychological atmosphere of silence is an example for phonetic symbolism in a poetic context. To substantiate this, he has no recourse to feature analysis, but makes a re-writing exercise:

2. When to the crux of crucial quiet thought
   I crave and call remembrance of things past.

Hrushovski points out that the [k] does not have the same hushing effect as the sibilants; and the focus is most likely shifted from sweet silent to crux. In the current study, we examined how the change between sibilant and [k] alliteration affects the emotive quality of the text. In addition, we note that although both excerpts are written in iambic pentameter, Excerpt 1 deviates from this structure in sweet silent. Thus, we wanted to test whether metric deviations and alliteration patterns can focus the reader’s attention to a specific phrase. For that purpose, we further altered the first line:

3. When to the quorum of kind quiet thought
   I crave and call remembrance of things past.

Excerpt 3 shares the [k] alliteration with Excerpt 2, but it deviates from a perfect iamb, like Excerpt 1 (the phrase kind quiet contains two consecutive stressed syllables). In addition, Excerpt 3 is semantically more similar to Excerpt 1 (e.g. quorum restores the legal connotations of “to the sessions … summon”). We presented participants with the three excerpts and asked them to evaluate the salience of phrases, the emotive quality of the texts, and the contribution of speech sounds to both aspects. The focus of this article is on the interaction between sound and emotive quality. We discuss the issue of salience in online Appendix IV.
We hypothesised that changing the alliteration pattern and the metric structure would affect the emotive quality of the text. In particular, we expected the emotive impressions evoked by Excerpt 1 to reflect the hushing, fluid (soft) quality of the sibilants that is further enforced by the meaning of sweet and silent and the weakened gestalt caused by the metric deviation. By contrast, [k] is abrupt, plosive, highly-encoded, pronounced by focused energy, and has an explosive and hard potential. Thus, it does not become expressive of some “kind, quiet” atmosphere originating in the meaning of the words in Excerpts 2 and 3. We, therefore, may expect that the readers will impute something strong and harsh to the latter two excerpts. Moreover, it is possible that the readers will perceive a conflict between the hard [k] sound and meaning of quiet and kind, or find no relation at all between sound and meaning in these excerpts.

Methods

Participants

Fifty-three participants completed the survey (age: mean: 40, range: 19–78; 32 women). All but two were native speakers of English and were residents of several countries, mainly the United States, Canada, and Israel. They had variable background in English poetry, from novice (high school or introductory university level) to expert (professors of literature specialising in English poetry). The distribution of literary experience of the participants is shown in Table 1.

<table>
<thead>
<tr>
<th>Level of literary training</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>No formal training at university level</td>
<td>1</td>
</tr>
<tr>
<td>Basic university training (BA students majoring in English literature, English teachers with no degree in literature)</td>
<td>9</td>
</tr>
<tr>
<td>BA level &amp; MA students</td>
<td>18</td>
</tr>
</tbody>
</table>

 Manipulation of the sound variables necessarily entailed unplanned semantic changes too.

 Since we wanted informed responses to elusive literary issues we recruited participants from variable backgrounds. Despite our initial hypothesis that participants with more extensive background in literature might be more sensitive to sound-emotion effects, the results of the experiment did not provide conclusive evidence for any education-related influence.
MA level & PhD students | 11
---|---
PhD | 14

Participants were recruited via university mailing lists and the PSYART forum. They were awarded an electronic gift card on the amount of 20£ for Amazon.8

Procedure

The experiment was run on the Qualtrics platform (https://www.qualtrics.com). Participants were asked to read Excerpts 1-3 out loud and answer several questions about each of them. All questions were mandatory, unless stated otherwise. The original text (Excerpt 1) was presented first, followed by some questions:

Read the following lines out loud:

*When to the sessions of sweet silent thought*

*I summon up remembrance of things past*

A) Do these lines evoke certain emotions or feelings? (e.g., happiness, anger, sadness, etc.) Specify any terms that come into mind.

B) Consider each of the following terms, and rate how well it reflects the emotional effect of the lines. (1 star = the term does not reflect the emotional effect of the lines, 5 stars = the term highly reflects the emotional effect of the lines):

The terms for evaluation for question (B) were: *hushed, intimidating, reflective, aggressive, grief*. A screenshot displaying question (B) is shown in Figure 1.

---

8 The current study was part of a larger study on the perceived qualities of sound patterns in poetry. Another part of that study was published elsewhere (Tsur & Gafni, 2018).
Consider each of the following terms, and rate how well it reflects the emotional effect of the lines. (1 star = the term does not reflect the emotional effect of the lines, 5 stars = the term highly reflects the emotional effect of the lines):

<table>
<thead>
<tr>
<th>Term</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>relaxing</td>
<td>⭐⭐⭐⭐⭐ 5</td>
</tr>
<tr>
<td>hushed</td>
<td>⭐⭐⭐⭐⭐ 5</td>
</tr>
<tr>
<td>intimidating</td>
<td>⭐⭐⭐⭐⭐ 1</td>
</tr>
<tr>
<td>reflective</td>
<td>⭐⭐⭐⭐⭐ 4</td>
</tr>
<tr>
<td>aggressive</td>
<td>⭐⭐⭐⭐⭐ 1</td>
</tr>
<tr>
<td>grief</td>
<td>⭐⭐⭐⭐⭐ 3</td>
</tr>
</tbody>
</table>

**Figure 1. Statement rating questions.** The scales were displayed to the participants without any selected values. The values selected in the figure are shown for the purpose of demonstration.

At this point, the excerpt was presented again, this time with certain letters in boldface, followed by another two questions:

*When to the sessions of sweet silent thought*

*I summon up remembrance of things past*

C) Consider the sounds of the boldface letters. Do these sounds contribute to the emotional effect you described in the previous question? [participants were expected to choose the option that reflected their impressions]

- Yes, I feel that the sounds help generating this emotional effect.
- No, I don't feel a direct connection between the sounds and the emotional effect of the lines.
- No, the sounds actually evoke a different emotional effect than the effect of the lines.

D) If you feel that the emotional effect evoked by the sounds of the boldface letters is different from the general effect of the lines, please elaborate on this difference. [filling-in this evaluation was not mandatory]
After completing the evaluation of Excerpt 1, participants performed the same task on Excerpts 2 and 3 (in that order). We describe only the non-identical instructions for these excerpts. Excerpt 2:

Read the following lines out loud:

*When to the crux of crucial quiet thought
I crave and call remembrance of things past*

Second presentation of the text, with some letters in boldface (followed by questions (C) and (D) as before):

*When to the crux of crucial quiet thought
I crave and call remembrance of things past*

Excerpt 3:

Read the following lines out loud:

*When to the quorum of kind, quiet thought
I crave and call remembrance of things past*

Second presentation of the text, with some letters in boldface (followed by questions (C) and (D) as before):

*When to the quorum of kind, quiet thought
I crave and call remembrance of things past*

After evaluating the three excerpts, individually, the participants were asked to compare the excerpts in free style [Filling-in this evaluation was not mandatory]:

E) Please make any comments you may have on a comparison between the three excerpts on this page.

In addition to the above tasks, participants were also given the opportunity to leave general comments on the experiment. We hoped that such comments would provide more detailed information about the participants’ thought processes.
Analysis and Results

The aim of the present study was to test various factors that might affect the emotive quality of poetic text. For that purpose, we conducted a comparative analysis of the responses to the three excerpts.

Emotive effect

We begin by analysing the open-ended emotive impressions (question A). We tokenised all the responses, and went over them manually to extract all the words that referred directly to the perceived effect of the text (hereinafter, keywords). For example, from the response “It brings up emotions of nostalgia and somberness”, we extracted the words “nostalgia” and “somberness” as keywords. Overall, 117 different keywords were specified by the participants in the descriptions of the three excerpts. Table 2 lists keywords used by five or more participants (see full lists in online Appendix I).

Table 2. Frequent terms describing the emotive effects of the excerpts

<table>
<thead>
<tr>
<th>Excerpt 1 (“Sessions”)</th>
<th>Excerpt 2 (“Crux”)</th>
<th>Excerpt 3 (“Quorum”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word</td>
<td>N</td>
<td>Word</td>
</tr>
<tr>
<td>nostalgia</td>
<td>21</td>
<td>longing</td>
</tr>
<tr>
<td>sadness</td>
<td>9</td>
<td>sadness</td>
</tr>
<tr>
<td>happiness</td>
<td>6</td>
<td>anger</td>
</tr>
<tr>
<td>bittersweet</td>
<td>5</td>
<td>desire</td>
</tr>
</tbody>
</table>

N = the number of participants mentioning each keyword in their open-ended comments.

The above table suggests some similarities as well as some differences among the excerpts. All three excerpts evoked feelings of sadness and yearning (nostalgia, longing). This is probably evoked by remembrance of things past. However, the feeling of nostalgia was much more dominant in the responses to Excerpt 1 than in the other two excerpts. Moreover, while Excerpt 1 and 3 evoked a dominant feeling of happiness, Excerpt 2 evoked more anger responses.

---

9 We should note that although it is possible that the instructions given to the participants biased them to use certain keywords more than others, the variety and distribution of keywords indicate that, if there was such a bias, it was rather marginal. Similarly, one could speculate that the responses were skewed by a retrieval effect, such that participants would tend to respond to subsequent questions using the same keywords. Again, the distribution of keywords across responses to the different texts suggests that this is unlikely to be the case.
To test whether the selection of keywords reflects genuinely different emotive effects, we mapped the keywords onto a three-dimensional emotion space whose axes are Valence, Arousal and Dominance. This allowed us to test, quantitatively, whether responses to the three excerpts distribute differently over the emotion space. Mapping of keywords was done using values of valence, arousal and dominance extracted for each keyword from The online NRC Valence, Arousal, and Dominance (VAD) Lexicon (Mohammad, 2018). These values are normalised judgments ranging from 0 to 1 (the average of the entire lexicon on each dimension is 0.5). For example, the valence, arousal, and dominance values for *nostalgia* are 0.417, 0.337, and 0.339, respectively; all of which are somewhat below average. In other words, *nostalgia* is somewhat negative, low-aroused, and low dominance. Out of the 117 keywords specified by the participants, 103 had an exact match in the NRC VAD Lexicon. For the other 14 keywords, we used the values for a (near-) synonym or the average values over several (near-) synonyms (e.g. for the word *soothed*, which was not found in the lexicon, we used the average values for *soothe* and *soothing*). The mean emotive values of the impressions on each dimension are visualised in Figure 2 (A – valence, B – arousal, C – dominance).
Figure 2. Quantitative emotive judgments. Mean emotive values of the impressions collected for the three excerpts are shown on three dimensions: (A) Valence, (B) Arousal, and (C) Dominance. Error bars indicate ±1 standard error of the mean.

* $p < .05$, ** $p < .01$, *** $p < .001$

To determine whether the three excerpts have different emotive effects, we conducted one-way multivariate analysis of variance (MANOVA) with valence, arousal and dominance as dependent variables and excerpt as an independent variable.\(^{10}\)

\(^{10}\) We should emphasise that we do not regard the emotive values of the keywords as characterising the responses themselves but rather the object to which the responses refer (i.e., the texts).
The analysis showed a significant difference among the excerpts (Wilks’ lambda = .72, F(6, 316) = 9.53, p < .001, partial $\eta^2 = .15$). Separate analyses of variance conducted for single dependent variables revealed significant differences across excerpts with respect to all three dimensions (Valence: F(2, 160) = 5.18, p < .01, partial $\eta^2 = .06$; Arousal: F(2, 160) = 23.86, p < .001, partial $\eta^2 = .23$; Dominance F(2, 160) = 3.61, p < .05, partial $\eta^2 = .04$).

Pairwise comparisons revealed that, on the Valence dimension, Excerpt 2 (“crux”; M = .38, SD = .04) evoked significantly more emotionally negative responses than Excerpt 1 (“sessions”; M = .54, SD = .04, p < .01) and Excerpt 3 (“quorum”; M = .53, SD = .05, p < .05). In addition, Excerpt 2 (M = .65, SD = .03) evoked more aroused responses than Excerpt 1 (M = .38, SD = .03, p < .001) and Excerpt 3 (M = .43, SD = .03, p < .001). Finally, responses to Excerpt 2 (M = .52, SD = .03) were characterised by a higher degree of Dominance compared to responses to Excerpt 1 (M = .42, SD = .03, p < .01), but not compared to Excerpt 3 (M = .47, SD = .03, p = .23). There were no significant differences between Excerpt 1 and Excerpt 3 on either dimension.

Overall, the analysis suggests that the emotive impressions were significantly influenced by semantics. Excerpt 2 evoked more negative and aroused responses, likely due to the influence of *crux* (and possibly also *crucial*), while Excerpts 1 and 3, which are semantically similar, evoked responses of similar nature. We should note, however, that the nostalgic quality is much more pronounced in Excerpt 1 than Excerpt 3 (and Excerpt 2). This could suggest some effect of sound pattern, as well: in short, the hushing sibilants in Excerpt 1 are more compatible with the low arousal state of nostalgia, and thus, amplify the sense of nostalgia. By contrast, [k] in Excerpt 3 weakens the sense of nostalgia, while [kr] and the semantics of *crux* and *crucial* in Excerpt 2 override it. We elaborate more on the contribution of sound patterns to the emotive effects in the next section.

While the open-ended descriptions collected in question (A) provided us a glimpse into the free impressions of the participants, the rated statements in question (B) allowed us to test some more specific hypotheses concerning the emotive effects of the excerpts based on the expected semantic and phonetic effects. The examined emotions included *hushed* and *reflective* that are particularly relevant to Excerpt 1, based on both semantic and phonetic factors. These terms are compatible with Excerpt 3 mainly in

11 Statistical analysis was performed on SPSS 21 (IBM Corp, 2012).
semantics, and are hardly appropriate in the context of Excerpt 2. To that, we added *relaxing* and *grief*, another two low-arousal emotional states, and also *intimidating* and *aggressive*, which are high-arousal states. The latter two were expected to be more relevant to Excerpt 2. *Relaxing* is semantically similar to *hushed*, but is less associated with sibilant alliteration. *Grief* is not particularly compatible with any of the excerpts.

Table 3 compares the evaluation of pre-defined keywords across the three excerpts. For each keyword, the table indicates the median and mean ratings for each excerpt. We also report the results of One-sample Wilcoxon Signed Rank Tests (test statistic T* and p value), used for testing the hypothesis that participants had an opinion about each of the statements. We operationalised the concept of “having an opinion” as a median rating significantly different from 3.12 In addition, we report results from Friedman’s Two-way analysis of variance comparing the ratings of the various excerpts for each keyword. Pairwise comparisons were done using Dunn’s test.

Table 3. Rated emotive judgments

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Excerpt 1 (“Sessions”)</th>
<th>Excerpt 2 (“Crux”)</th>
<th>Excerpt 3 (“Quorum”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>relaxing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Mean</td>
<td>3.08</td>
<td>1.57</td>
<td>3.30</td>
</tr>
<tr>
<td>T* (median ≠ 3)</td>
<td>0.53</td>
<td>-5.56</td>
<td>1.79</td>
</tr>
<tr>
<td>p</td>
<td>.59</td>
<td>&lt;.001 ***</td>
<td>.074</td>
</tr>
<tr>
<td>hushed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Mean</td>
<td>3.74</td>
<td>1.62</td>
<td>2.85</td>
</tr>
<tr>
<td>T* (median ≠ 3)</td>
<td>4.18</td>
<td>-5.41</td>
<td>-0.68</td>
</tr>
<tr>
<td>p</td>
<td>&lt;.001 ***</td>
<td>&lt;.001 ***</td>
<td>.50</td>
</tr>
<tr>
<td>reflective</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Mean</td>
<td>4.58</td>
<td>2.72</td>
<td>3.51</td>
</tr>
</tbody>
</table>

12 Since the marks on the evaluation scales indicated degrees of agreement with statements, we assume that if participants did not understand a statement or were unable to judge it, they would give it an intermediate rating, namely 3 on a scale of 1-to-5.
<table>
<thead>
<tr>
<th></th>
<th>Median</th>
<th>Mean</th>
<th>T* (median ≠ 3)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>intimidating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.26</td>
<td>2.83</td>
<td>1.42</td>
<td></td>
</tr>
<tr>
<td>T* (median ≠ 3)</td>
<td>-6.59</td>
<td>-1.40</td>
<td>-6.40</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>&lt;.001 ***</td>
<td>.15</td>
<td>.007 **</td>
<td></td>
</tr>
<tr>
<td><strong>aggressive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.25</td>
<td>3.53</td>
<td>1.40</td>
<td></td>
</tr>
<tr>
<td>T* (median ≠ 3)</td>
<td>-6.51</td>
<td>2.61</td>
<td>-6.43</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>&lt;.001 ***</td>
<td>.009 **</td>
<td>&lt;.001 ***</td>
<td></td>
</tr>
<tr>
<td><strong>grief</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.66</td>
<td>2.32</td>
<td>2.02</td>
<td></td>
</tr>
<tr>
<td>T* (median ≠ 3)</td>
<td>-1.91</td>
<td>-3.33</td>
<td>-4.33</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>.06</td>
<td>.001 ***</td>
<td>&lt;.001 ***</td>
<td></td>
</tr>
</tbody>
</table>

T* = standardised One-sample Wilcoxon Signed Rank Test statistics. Used for testing whether participants tended to agree/disagree with a statement as opposed to having no opinion (a lack of opinion was operationalised as median rating of 3, on a scale of 1-to-5).

** p < .01, *** p < .001

Table 3 suggests that participants had no opinion about whether Excerpts 1 and 3 had a relaxing effect. However, they strongly disagreed that Excerpt 2 was relaxing. A two-way Friedman’s ANOVA confirmed that the three excerpts were not relaxing to a similar degree ($\chi^2(2) = 48.50, p < .001$). Pairwise comparisons revealed that this difference is due to Excerpt 2 being judged as less relaxing than Excerpt 1 ($z = 5.1, p < .001$) and Excerpt 3 ($z = 5.97, p < .001$). The analysis of hushed indicates that participants discerned such an effect in Excerpt 1, were of no conclusive opinion regarding Excerpt 3, and showed strong disagreement regarding Excerpt 2. This three-way difference was supported by a two-way Friedman’s ANOVA ($\chi^2(2) = 54.53, p < .001$). Pairwise comparisons confirmed that Excerpt 2 had a less hushing effect than
Excerpt 1 ($z = 6.67, p < .001$) and Excerpt 3 ($z = 4.36, p < .001$). The difference between Excerpts 1 and 3 was only near significant ($z = 2.30, p = .06$).

The results for the low-arousal states are expected based on the combination of semantic and phonetic effects. Excerpt 2 contains the words *crux*, *crucial*, and *crave* that convey a sense of unrelaxed urgency. This is probably enforced by the [k] (or even [kr]) alliteration, as some participants noted (e.g. “the alliteration and consonance of the "C" sound […] creates a harsh sound that corresponds with the denotative weight of the words "crux," "crucial," and "crave."”). Although Excerpt 2 contains the word *quiet*, which is compatible with relaxing and hushed qualities, it seems that the overall atmosphere was dominated by *crux* and *crucial*.

The [k] alliteration is also present in Excerpt 3. The fact that this excerpt was not judged as particularly relaxing or hushed could be attributed to the hard quality of [k], which many participants acknowledged in their verbal comments (e.g. “the harsh "K" or "C" sounds don't align with the denotations of the words "crave," "quiet," or "kind." They evoke a harsher, more negative sound, not a relaxing or quiet or hushed feeling.”). The formal/unfamiliar meaning of *quorum* might have also played a role here.

Yet, Excerpt 3 was judged as more relaxing and hushed than Excerpt 2. In addition, Excerpt 3 was not judged as less relaxing than Excerpt 1 ($z = -.87, p = 1$), contra to our expectation. One possible explanation for these results is that the hard potential of the [k] alliteration was softened by strong semantic effects (e.g. due to *kind* and *quiet*). Another explanation is that the hard potential of [k] was softened by the following [w] in *quorum* and *quiet* ([w] is a periodic, lowly-encoded, semivowel). This, in fact, was suggested by several participants (e.g. “This time the sounds are twice followed by a "w" sound that softens them (quorum, quiet)”). Thus, we may say that the emotive potential of [kw] is ambiguous. “Crucial” and “kind”, as well as regular and irregular meter, foreground different aspects of it.

As for Excerpt 1, semantics (*sweet silent*) and phonetics (sibilant alliteration), as well as the metric gestalt blurred by the consecutive stresses, work together to support a hushing effect, as expected. The more intriguing find was that Excerpt 1 was not judged as particularly relaxing. This could be due to the low valence (e.g. melancholic) associations that many participants attributed to Excerpt 1, which are incompatible with the high valence *relaxing*. Anyway, the different results for *hushed* and *relaxing* can be used as evidence that the general categories, such as *sad* and *joyful*, typically used in
empirical studies of poetry provide only a crude estimation of the poem’s emotive quality.

Deviation from metre may also support the observed effects. The phrase *sweet silent* in Excerpt 1 contains two adjacent stressed syllables, the first of which (*sweet*) is in a metrical weak position. Similarly *kind* in Excerpt 3 also contains a stressed syllable in a weak position. Such metric deviations soften gestalts and may support low-energy emotions (though there are other possibilities, as we mention later).

The three excerpts also differed in the degree of reflective impressions ($\chi^2(2) = 45.96, p < .001$). Excerpts 1 and 3 were judged as reflective, while participants were neutral about Excerpt 2. The differences between Excerpt 1 and Excerpt 2 ($z = 5.93, p < .001$) and between Excerpt 1 and Excerpt 3 ($z = 3.55, p = .001$) were significant. The difference between Excerpt 2 and 3 was only near significant ($z = 2.38, p = .05$). The results were expected based on semantic and phonetic effects. *Remembrance of things past* and *thought* endowed all the excerpts with some reflective quality. In Excerpts 1 and 3 this is further supported by the semantics of *silent*, *kind*, and *quiet* which, like reflective, denote low arousal states. Although Excerpt 2 also contains *quiet* and *thought*, their effect is counteracted by the more aroused *crux* and *crucial* (though, note that participants did not completely reject Excerpt 2 as non-reflective). A potential phonetic effect might also be observed in the participants’ responses. As discussed above, metric deviations in Excerpt 1 and 3 may strengthen the effect of low arousal emotions. In addition, the fact that Excerpt 1 was judged as more reflective than Excerpt 3 could be attributed to the difference between sibilant alliteration and [k] alliteration. The soft quality of the sibilants reinforces low-energy emotions (a hushing effect), while [k] does not support them.

Next, participants rejected Excerpts 1 and 3 as intimidating, while they had no conclusive opinion about Excerpt 2. A two-way Friedman’s ANOVA confirmed that the three excerpts were not intimidating to a similar degree ($\chi^2(2) = 59.55, p < .001$). Pairwise comparisons revealed that this difference is due to Excerpt 2 being judged as more intimidating than Excerpt 1 ($z = 5.63, p < .001$) and Excerpt 3 ($z = 4.71, p < .001$). Similarly, participants strongly disagreed that Excerpts 1 and 3 were aggressive, while they tended to agree that Excerpt 2 was aggressive. The three-way distinction was significant ($\chi^2(2) = 74.66, p < .001$), and so were the differences between Excerpts 2 and 1 ($z = 6.75, p < .001$) and between Excerpts 2 and 3 ($z = 5.93, p < .001$). The difference between Excerpts 1 and 3 was not significant ($z = .83, p = 1$).
The results for Excerpt 2 are likely due to combined semantic and phonetic effects. The “hard” [kr] alliteration resonates with the high-arousal high-dominance crucia_l and with crux, which has unpleasant connotations (it could be speculated that [k], or even the [kr] sequence, transfers its hard potential to the meaning of crucial). Note that no word in Excerpt 2 is particularly intimidating and, thus, the inconclusive judgment, but the overall effect is compatible with an intimidating atmosphere. By contrast, the semantics of Excerpts 1 and 3 clearly reject feelings of aggression and intimidation. In the case of Excerpt 3, the hard potential of [k] remains unexploited (what Hrushovski calls “neutral sound patterns”). One may also add that [kw] is softer than [kr].

Finally, impressions of grief also differed across the three excerpts ($\chi^2(2) = 12.5, p = .002$), though the effect was considerably weaker compared to the previous evaluations. Participants tended to disagree that Excerpts 2 and 3 evoke feelings of grief, but had no clear opinion regarding Excerpt 1. The only significant pairwise comparison was between Excerpts 1 and 3 ($z = 2.91, p = .01$). A priori, all three excerpts could evoke impressions of grief due to the potentially sad connotations of remembrance of things past. However, in Excerpts 2 and 3, the high-arousal semantics of crux and crucial and the hard [k] alliteration seem to override the low-arousal grief. By contrast, in Excerpt 1, the low-arousal atmosphere generated by the semantics and supported by the sibilant alliteration is more compatible with feelings of grief. Here, the effect is likely mitigated due to the conflict between the negative grief and the positive sweet.

Sound-emotion interaction

In this section, we further investigate the contribution of sound patterns to the impressions evoked by the excerpts. Recall that after answering several questions about each excerpt, participants were shown the excerpt again, this time with certain letters in boldface. The boldfaced letters represented sibilant consonants in Excerpt 1 and [k] consonants in Excerpt 2 and 3. The participants were then asked about the connection between the sound pattern and the emotive effect of each excerpt (question C: “Consider the sounds of the boldface letters. Do these sounds contribute to the emotional effect you described in the previous question?”). Table 4 summarises the responses to question (C).
In general, participants agreed that sound patterns contributed to the emotive effects of the three excerpts. A Chi-square test did not find a significant difference among the excerpts (\(\chi^2(4) = 6.44, p = .17\)), though it could be observed that participants tended to find less connection between the sound pattern and the emotive effect in Excerpt 3. According to our hypothesis, the connection between sound and emotive quality should be the strongest in Excerpt 1, in which the sibilants enforce the meaning of the words. In Excerpt 2 we expected more variable responses due to the conflict between the hard [k] and quiet (although, [k] was somewhat compatible with crux and crucial). Finally, the sound-emotive quality connection was expected to be the weakest in Excerpt 3 due to the conflict of [k] with kind and quiet. These expectations were only partially confirmed.

To further investigate the relation between sound and emotive effect, we examined the detailed responses to question (D) (If you feel that the emotional effect evoked by the sounds of the boldface letters is different from the general effect of the lines, please elaborate on this difference; see full lists of comments in online Appendix II). In addition, we examined comments related to the sound content that were made in response to the final question, which asked for direct comparison of the three excerpts (question E; see full lists of comments in online Appendix III).

Many participants provided informative comments on the effect of the repeated sibilants. Most of them indicated their hushing, relaxing effect in Excerpt 1 (e.g. “The

<table>
<thead>
<tr>
<th>Response</th>
<th>Excerpt 1 (“Sessions”)</th>
<th>Excerpt 2 (“Crux”)</th>
<th>Excerpt 3 (“Quorum”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>41</td>
<td>42</td>
<td>33</td>
</tr>
<tr>
<td>No direct connection</td>
<td>9</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>The sounds evoke a different emotional effect</td>
<td>3</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

In general, participants agreed that sound patterns contributed to the emotive effects of the three excerpts. A Chi-square test did not find a significant difference among the excerpts (\(\chi^2(4) = 6.44, p = .17\)), though it could be observed that participants tended to find less connection between the sound pattern and the emotive effect in Excerpt 3. According to our hypothesis, the connection between sound and emotive quality should be the strongest in Excerpt 1, in which the sibilants enforce the meaning of the words. In Excerpt 2 we expected more variable responses due to the conflict between the hard [k] and quiet (although, [k] was somewhat compatible with crux and crucial). Finally, the sound-emotive quality connection was expected to be the weakest in Excerpt 3 due to the conflict of [k] with kind and quiet. These expectations were only partially confirmed.

To further investigate the relation between sound and emotive effect, we examined the detailed responses to question (D) (If you feel that the emotional effect evoked by the sounds of the boldface letters is different from the general effect of the lines, please elaborate on this difference; see full lists of comments in online Appendix II). In addition, we examined comments related to the sound content that were made in response to the final question, which asked for direct comparison of the three excerpts (question E; see full lists of comments in online Appendix III).

Many participants provided informative comments on the effect of the repeated sibilants. Most of them indicated their hushing, relaxing effect in Excerpt 1 (e.g. “The

---

13 We did a second analysis that included responses from participants who did not complete the experiment (63, 60, and 59 responses to Excerpt 1, 2, and 3, respectively). This analysis did reveal a significant difference due to a higher tendency of participants to find no connection between sound pattern and emotional effect in Excerpt 3 compared to Excerpt 2. Although this tendency is also implied by the final data, it did not reach statistical significance.

14 Although many of the responses suggested that the effect of the sound pattern was similar to the general effect of the excerpt (and, thus, require no further elaboration), they were informative, nonetheless.
alliterative S sound made the sentences seem calm, smooth and easing their way into my consciousness. No aggressive sounds.”). Participants also commented on the effect of the [k] alliteration in Excerpt 2. They indicated the harsh, aggressive quality of [k]. Some of them emphasised the conflict with the notion of quiet thought. One participant suggested that the alliteration was, in fact, compatible with the semantics of the words crux, crucial, and crave (“the alliteration and consonance of the "C" sound […] creates a harsh sound that corresponds with the denotative weight of the words "crux," "crucial," and "crave.").

The harsh quality of [k] was also mentioned by several participants in the context of Excerpt 3. Again, participants noted the conflict between the perceptual quality of the consonant and the relaxed quality suggested by the semantics. One participant commented that the metric deviation in kind quiet (a sequence of two stressed syllables) reinforced the conflicting effects (“The hard "K" sounds conflict with the hushed, reflective idea of "quiet thought." The fact that two stresses appear without an intervening unstressed syllable (e.g. "kind, quiet") adds to the dissonance.”).

Finally, although the instructions focused on the emotive effects of sibilant consonants and [k], some participants noticed the interaction with other consonants. In Excerpt 2, the combinations of [k] and [r] in crux, crucial, and crave seemed to magnify the harshness. By contrast, the glide [w] seemed to soften the preceding hard [k] in quorum and quiet in Excerpt 3. One participant in particular noticed the contrast between the sound combinations in the two excerpts: “I think that it is a combination of letters. Where the "kr" combination of sounds felt very harsh to me, the "kw" sound is calmer and sadder.”

**Discussion**

In this paper, we proposed a method to quantify open-ended responses to poetry. We argue that this method is particularly suitable for studying the subjective experience evoked by aesthetic objects, since it does not confine judgments to researcher’s predefined dimensions. We combined this method with a qualitative analysis of verbal comments and quantitative analysis of close-ended questions to investigate how sound patterns, the meaning of words, and prosody interact to affect the emotive quality of poetic text. Such combination of methods can provide information not only about the
bottom-line preferences of participants, but also about the process of reaching the preferences. We illustrate this claim below.

The three excerpts used in the experiment share the common theme of quiet reminiscing. Hence, it is not surprising that words like *nostalgia* and *longing* were common in participants’ responses to all of them. However, the nostalgic atmosphere is “coloured” by different emotions in each excerpt. In Excerpts 1 and 3, the perceived quality is of passive, bittersweet nostalgia (intermediate valence, low arousal). This is reflected by intermediate-to-high ratings of the excerpts along the *relaxing*, *reflective* and *hushed* scales, as well as frequent use of *happiness* and *sadness* as free descriptive terms. By contrast, in Excerpt 2 the memories evoke a sense of anger and urgency (low valence, high arousal). This is reflected by intermediate-to-high ratings of the excerpt along the *aggressive* and *intimidating* scales, as well as frequent use of *anger* and *desire* as free descriptive terms. Such subtle distinctions are hard to detect with the common methodology, which relies exclusively on responses according to predefined qualities, like grief–joy or tender–aggressive. Moreover, the free impressions may reveal additional aspects of the aesthetic event that have escaped the researcher’s attention. For example, the verbal responses suggest that one of the causes for the differential effects of Excerpts 2 and 3 was the contrast between the [kr] and [kw] alliteration patterns ([r] enhanced the effect of [k] alliteration while [w] mitigated it). We did not anticipate such interactive effects, though they seem rather sensible in retrospect (see more below).

While the current study is novel in its approach, it is not the first to use quantitative dimensional models of emotion in empirical studies of poetry. Some researchers (e.g. Aryani et al., 2016; Ullrich et al., 2017; Whissell, 1999) use external affective dictionaries to predict the emotive quality of poems based on the emotive values (e.g. valence, arousal) of the words of the poem.\(^{15}\) However, we claim that the emotive quality of a poem cannot be determined this way as straightforwardly as in plain descriptive texts. One reason for this is that the context in the poem can change the emotive connotations of words. For example, the word *moan* may have negative

\(^{15}\) Aryani et al. (2016), Ullrich et al. (2017), and Whissell (1999) go one step further and attempt to predict the emotional qualities of poems based on their sound contents, assuming the emotive character of speech sounds can be derived from the meanings of words in which they occur. However, this approach assumes that the relationship between speech sounds and meanings is non-arbitrary throughout the dictionary, which is usually not the case.
connotations in some contexts, but in Tennyson’s “Come down, O Maid” *The moan of doves* is used metaphorically to describe sounds of pleasure (Tsur & Gafni, 2019a).

The current study also departs from previous studies in its approach to phonetic symbolism. Our account for the sound-emotion interactions in these excerpts relies on structural analysis of speech sounds and emotions and on transfer of features. Associations between speech sounds and emotions are formed on the basis of structural resemblance. Thus, it is not the case that the sibilants are intrinsically hushing, or that [k] is intrinsically harsh. Rather, the acoustic features of these sounds make them potentially suitable to support various emotional atmospheres. For example, “intimidating” and “aggressive” are characterised by high arousal, may involve sudden outbursts, and are typically focused on some target. Likewise, [k] is acoustically abrupt, a plosive (that is, produced by energy transmitted outwards as a shock wave).\(^{16}\) Thus, the features of [k] make it suitable to support high arousal states such as *aggression*, but also *joy*. The sibilants, by contrast, consist of a stream of continuous, irregular noises, and may be characterised, perceptually, as fluid (that is, unfocused). Thus, they may both imitate noises and express softness and hushing. *Silent* picks out the hushing potential which, in turn foregrounds the hushing quality of the excerpt.

Speech sounds support the general emotional atmosphere not only by resonating with word meanings they share features with, but also by activating similar features in other words whose meaning does not cancel their effect. Thus, [k] is prone to transfer, when prompted, features shared with the above high-arousal emotions to words whose meaning does not cancel their effect, like “crux” and “crucial”. Similarly, sibilants may transfer, when prompted, features shared with low-arousal emotions to words like “sweet”. Importantly, speech sounds are arbitrary signifiers and their emotive potential is essentially dormant unless meaning activates it. Thus, the [k] alliteration remains rather “neutral” in Excerpt 3, since the meanings of “kind” and “quiet” mitigate its effect. Although, the fact that, when explicitly asked, participants noticed a conflict between the semantics and the alliteration pattern suggests that the semantics does not cancel the alliteration effect altogether.

\(^{16}\) [k] shares this feature with the other voiceless plosives (e.g. [t] and [p]). It differs only in being more focused, which makes it an exceptionally hard plosive. In addition, focused energy may foreground the focused quality of aggression. This may perhaps explain, why Hrushovski (1980) chose [k] out of all plosives for his exercise.
Interestingly, interaction can be not only between sounds and meaning, but also between different sounds in the same text. In fact, as some participants suggest, the different effects of Excerpt 2 and 3 could be partly due to the interaction of [k] with [r] and [w]. [w] is a labialised periodic consonant, and these features may support a softened atmosphere (Gafni & Tsur, 2019). The fact that [r] is perceived as “hard” is more difficult to account for. We hypothesise that this is due to its spectral properties.\textsuperscript{17}

In the present study, we were also interested in the effect of metric deviation. “Sweet silent” and “kind quiet” deviate from iambic metre in contrast to “crucial quiet”. Metric deviations may have different effects on readers, depending on whether the reader is able to accommodate the deviance in a rhythmic performance. This reflects the rhythmic competence or lack of competence even of professors. In this regard, we distinguish four degrees of rhythmic competence in an ascending order of aesthetic complexity: 1. metric deviation is ignored; 2. metric deviation disturbs metre, rendering it harsher; 3. metric deviation draws attention to the words involved, foregounding their meanings; and 4. metric deviation is accommodated in a rhythmical performance, rendering its gestalt softer.

We have received explicit comments on regular and deviating meter from only three participants, and for them, the effect seems to be of disturbance (see more in Appendix IV). We cannot determine whether other participants noticed these deviations or how they perceived them. It is, however, possible that some participants spontaneously performed the deviant verses rhythmically and were implicitly affected by rhythm. Assuming such a possibility, how would rhythm affect the perceived difference between Excerpts 1 and 3 on the one hand, and Excerpt 2 on the other?

The regular alternation of stresses in “of crúcial quiet thóught” is perceived as strong gestalt; the consecutive stresses in “of swéet sílent thóught” and “of kind quiet thóught” blur the gestalt of iambic meter. According to Leonard B. Meyer strong gestalts in music generate “a psychological atmosphere of certainty, security, and patent purpose, in which the listener feels a sense of control and power as well as a sense of specific tendency and definite direction” (Meyer, 1956: 160). Metric deviance in “sweet

\textsuperscript{17} Another possible (albeit less likely) account for the “hardness” of [r] in English is a backward influence from the lexicon. Until a certain unknown point in the past, the English [r] was a trilled consonant, namely, multiply interrupted (in most modern dialects of English [r] is an approximant). This property gives [r] a “hard” potential by the same mechanism that works in plosives. The hard potential may be detected in a large number of lexical items (such as “roar”, “hard”, “harsh”) which, in turn, may affect how the [r] is perceived by native speakers of English today.
silent” and “kind quiet”, if accommodated in a rhythmic performance, may drastically reduce such an effect. Participants found Excerpt 2 opposed to the other two excerpts on a number of dimensions; among others, in being more “dominant”, more “angry”, and less “reflective”. The sense of control and power of strong gestalts would reinforce dominance, the sense of power and specific tendency would reinforce anger found on other grounds in Excerpt 2. The Merriam-Webster Collegiate Dictionary defines the relevant meaning of “reflect” as “to think quietly and calmly”, suggesting “unhurried consideration of something recalled to the mind”. The reduced sense of power, control and patent purpose experienced in weak gestalts would reinforce the reflective quality in Excerpts 1 and 3. Thus, even if participants did not explicitly notice metric difference, it may have influenced their overall impression.

The fact that only three participants explicitly acknowledged the effect of metric deviations could be because these deviations were too subtle or because the task made participants focus on other aspects of the texts. Either way, the weak evidence for prosodic effects can serve as a reminder that responses to poetry reflect not only the properties of the text but also the participants’ competence and the properties of the experimental design.

We would like to conclude the discussion by pointing out some limitations in the current study and potential endeavours for future research. Most importantly, although the quantitative analysis of open-ended responses can potentially overcome the disadvantages of forced-choice tasks, it is, in itself, not foolproof, either. The free responses we analyse reflect the elusive intuitions of the participants, for which they try to find the best words. Consequently, they end up with different terms even when they are hunting down similar intuitions. Our methodology looks for similarities and differences across responses in order to obtain meaningful, quantitative measures. But, in doing so, it reduces the semantic variability of the responses, and consequently, provides only general description of the emotive quality of the text.

In addition, it is noteworthy that, while the concept behind our methodology is general, the concrete results depend on the emotive model used for the analysis of the responses. In particular, although the VAD model is widely used to map words into an emotive space, there might be other, more suitable models to capture the differences in
emotive qualities across poetic texts. Thus, we argue that a quantitative analysis of the free responses should be complemented by a qualitative discussion of the responses.\textsuperscript{18}

Another potential limitation of the current study lies in the fact that the experiment was done online, so we had little control over what the participants did. In particular, we instructed participants to read the excerpts out loud, and we have no way to know whether all of them followed the instruction. However, it is important to note that the purpose of this instruction was to facilitate the participants’ engagement with the texts, similarly to what literary critics do when they read a poem. We did not collect recordings from the participants since we were interested in how they perceived the texts and not in how they produced the texts. Future studies could benefit from analysing recordings of texts recited by the participants, especially in studies of poetic prosody. Such studies may also incorporate neuroimaging (e.g. Teng et al., 2020) to probe the neural mechanisms involved in reading poetry and listening to recited poetry. In particular, it would be interesting to examine whether the emotive values of explicit responses to poetry correlate with activations in certain brain areas.

Finally, in the current paper, we focused on the emotive effects associated with two phonemic categories: sibilants ([s] and [ş]) and [k]. The goal of our study was to demonstrate how sound-emotion effects can be investigated from a structural perspective, and how open-ended responses to these effects can be analysed to give a more complete view of the nature of these effects. Since our methods are based on feature analysis, they can be easily utilised to investigate sound-emotion effects with other phonemic categories and other texts.

Declaration of Conflicting Interests
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
This research was supported by The Israel Science Foundation (grant No. 228/11).

\textsuperscript{18} It should also be noted that the proposed method is useful for capturing differences in emotive qualities. Thus, analysing responses to texts that have similar emotive qualities will likely yield non-significant results.
References
IBMC Corp.


