Methodological Issues in the Study of Phonetic Symbolism

*** Author accepted manuscript ***

Full reference:
Methodological Issues in the Study of Phonetic Symbolism

Reuven Tsur\textsuperscript{a}, Chen Gafni\textsuperscript{b}

Author Note

\textsuperscript{a} Department of Literature, Tel Aviv University
\textsuperscript{b} Department of Computer Science, University of Haifa

Correspondence may be addressed to Reuven Tsur, email: tsurxx@tauex.tau.ac.il
Abstract

There is a growing research literature on phonetic symbolism in poetry, sometimes with incongruent results. Through a theoretical structural analysis we show that, (a) individual speech sounds have (sometimes conflicting) potentials to suggest elementary percepts, such as abruptness, hardness, smoothness; and (b) from these elementary percepts some general psychological atmosphere may be abstracted that may be individuated in specific emotions as ‘love,’ ‘joy,’ or ‘anger,’ by semantic feature-addition. This proposal can reconcile incongruous research results. Sound-symbolic lexical entries are governed by similar principles, but fossilized. Large-scale statistical investigation may reveal significant sound-symbolic effects only when the same phonemes repeat throughout the poem. They may, however, miss conspicuous local sound effects, revealed only by local analysis. Some sceptical conclusions in the research literature may be due to this phenomenon. The proposed method may account not only for statistical correlations, but also for the perception of a pervasive emotional atmosphere in a poem.

Keywords: sound symbolism, iconicity, poetry, emotion
In recent years, the literary study of the relationship between speech sounds and emotions has been dominated by a few research dogmas that ensure certain important results but, by the same token, block some other findings, no less important. This article argues against a cluster of views propounded in several prominent papers on this topic — by Wiseman and van Peer (2002), Auracher, Albers, Zhai, Gareeva, and Stavniychuk (2011), Kraxenberger and Menninghaus (2016), and Whissell (1999, 2004). We disagree with them on a wide range of theoretical and methodological issues. In what follows, we briefly summarise our main claim. In subsequent sections we will elaborate on each argument.

Our claims on the theoretical level are as follows: First, we argue that phonetic symbolism in literature is governed by linguistic creativity, generating an indefinite number of unforeseeable qualities. Second, such cases of symbolism-through-linguistic creativity override any fixed relationships determined by “an inherited hardwired connection between different domains of functional areas in the brain” (Auracher et al., 2011, p. 6). Third, symbolic relations, by themselves, are not inherited, but the propensity to create such relations is.¹ Fourth, the cues to phonetic symbolism are in the acoustic–phonetic information and, to some extent, in the articulatory information, rather than in the phonological information. Fifth, perceptual and emotional qualities are typically generated after the event, by an interaction between phonetic and semantic features in a specific text. Sixth, only a small part of the phenomenon can be accounted for by predisposition for specific sound–meaning combinations, if at all. Seventh, in different contexts, one phoneme group (e.g., plosives) may generate a variety of sometimes incongruent emotional or perceptual qualities (e.g., joy, aggression, resoluteness, or discreetness). Eighth, these effects are determined by the semantic features present and the phonetic/phonological contrast involved. Thus, for

¹ This is done by abstracting, comparing, and contrasting phonetic, semantic and visual features. Such operations underlie our linguistic competence in general, and other competences as well.
instance, the voiced plosive [b] is grouped in most experimental studies with other voiced and voiceless plosives and contrasted to liquids and nasals, whereas in the bouba/kiki experiments it is opposed to a voiceless plosive, and in the takete/baluma experiments it is grouped with a liquid and a nasal and opposed to voiceless plosives.

On the methodological level, we argue that quantitative-confirmatory research is extremely important, but this is only a small part of the story. There are crucial issues that cannot be handled in this way, and some qualitative-exploratory method is indispensable, such as thought experiments, case studies and hypotheses derived from linguistic, psychological, neurological and ethological research. Relevant information can be gathered on a much more fine-grained level than in Stimulus–Response (S–R) statistical studies. This fine-grained information is available in the stimulus structure and in our common-sense knowledge about perceptions and emotions, from which, by analytic methods, relying on empirical research in phonetics, psychology and other disciplines, one may reconstruct the structures mediating between stimulus and response (such a mediating structure can explain how an accumulation of plosives can generate both a joyful and an aggressive quality).

In an earlier stage of our research we focussed on phonetic symbolism as a literary phenomenon and accorded the benefit of doubt to the conception that onomatopoeia proper and certain lexical regularities may be governed by different principles. Eventually we came to the conclusion that these too are governed by the same principles as the literary phenomena, but underwent a further development: fossilization (for Poetic Conventions as Cognitive Fossils, see Tsur, 2017b). Something like this may have been realized by some other researchers as well; a recent article, “Tracking semantic change in fl- monomorphemes in the Oxford English Dictionary” (Smith, 2016) directs to Tsur (2006) for the underlying phonetic mechanism.
A welter of conflicting findings

There has been a general agreement in the linguistic literature, at least since Ferdinand De Saussure (1916), that the relation between sound and meaning is arbitrary by large. Yet, a growing body of studies suggests that certain, non-arbitrary relations can be formed. For example, it has been observed cross-linguistically that high front vowels (like /i/) are predominant in the expression of ‘smallness’ and that low back vowels (like /o/) are more typical in expressions of ‘bigness’ (Ohala, 1994). (In English, however, there are such pairs of synonyms as little=small, and large=big.) This approach does not apply in poetry where, far from challenging the arbitrary conception, arbitrary linguistic signs are arranged, under the infinite creativity of symbolic language, in unpredictable combinations to yield apparently non-arbitrary sound-symbolic expressions. Such unpredictable combinations may even override the relationships epitomized by Jakobson and Waugh (2002) as “inmost, natural similarity association between sound and meaning” (p. 182). As we shall see toward the end of this article, the theory proposed here may account for some pre-aesthetic lexical regularities too. Without understanding this, the field remains chaotic. Indeed, recent attempts to systematically investigate sound iconicity in literature have yielded a pandemonium of sometimes conflicting results.

Tsur (2012, pp. 271–274) pointed out a discrepancy between Wiseman and van Peer’s (2002) and Fonagy’s (1961) findings. The former found that nasals are correlated with grief, plosives with joy; the latter found that they are correlated with tenderness and aggression, respectively. In an intercultural study, Auracher, Albers, Zhai, Gareeva, and Stavniychuk (2011) confirmed Wiseman and van Peer’s (2002) findings, but also found that “poems that have a relatively high frequency of plosive sounds are more likely to express a pleasant mood with high activation, whereas a relatively high frequency of nasal sounds indicates an unpleasant mood with low activation. Moreover, these findings are universal” (p. 1).
Another consequence of Auracher et al.’s (2011) argument is the belief that one is able to make a priori predictions about a poem’s tone based on its sound content. Consider the following two excerpts:

[1] and I
Thy shepherd pipe, and sweet is every sound,
Sweeter thy voice, but every sound is sweet;
Myriads of rivulets hurrying thro' the lawn,
The moan of doves in immemorial elms,
And murmuring of innumerable bees

(Tennyson: Come down, O Maid)

You cataracts and hurricanoes, spout
Till you have drench’d our steeples, drown’d the cocks!

(King Lear, III.2)

According to Auracher et al. (2011), from the accumulation of nasals in the last two lines of Excerpt 1, we should predict that the passage will be perceived as displaying some negative emotion with low activation, like sadness; from the accumulation of plosives in Excerpt 2, as displaying some positive emotion with high activation, like joy. As we shall see later, there is no certainty at all that Excerpt 1 displays grief, and, obviously, Excerpt 2 suggests anger rather than joy. In other words, the valence of a poetic text cannot be predicted based on its sound content alone.

In another study, Miall (2001) analysed passages from Milton’s Paradise Lost. He found that the passage concerning Hell (negative mood) contained a higher frequency of ‘hard’
consonants (i.e., plosives) than the passage referring to Eden (positive mood). In addition, he found that in Coleridge’s “Midnight Frost,” “glides, liquids, and nasals are more frequent in the positive lines” (Miall, 2001, p. 66).

Chaos reigns in findings concerning vowels too. And here, too, the remedy is taking the underlying mechanisms into consideration. Based on the findings of Wiseman and van Peer (2002), one would expect Milton to prefer in the description of hell dark vowels, bright vowels in the description of heaven. Miall (2001), however, found that Milton preferred ‘narrow’ front vowels in the description of hell, suggesting the confinement of hell, and wide back vowels in the description of heaven (p. 64). This manipulates ‘bright’ vowels into the hell passage, and ‘dark’ vowels into the heaven passage. In other words, Milton had recourse here to articulatory rather than acoustic features as the critical features. One may, of course, reject Miall’s interpretation. The alternatives to this would be that Milton’s use of vowel symbolism is anomalous, or non-existent.

Kraxenberger and Menninghaus (2016) tried to replicate Auracher et al.’s (2010) findings, but they failed:

The poem with the highest relational frequency of plosives was rated as sad and not, as would be expected based on previous findings, as joyful. At the same time, the poem with the highest frequency of nasals was also rated as sad. Thus, the results of Auracher et al. (2010) could not be replicated. Furthermore, joyful poems did not differ from sad poems in terms of relational or normalized frequencies of occurrence of plosives and nasals. Consequently, an iconic relation between these phoneme classes and emotional classification could not be confirmed. (p. 7)

Thus, the starting point of the present study is the existence of two lines of conflicting findings: one concerning the types of emotional qualities that can be expressed by the various sound classes; the second conflict is about whether individual sounds can have emotional
qualities at all. Since most studies seem to agree that some kind of iconic sound–emotion relation can be found in poems, we will start with a structural theory of sound–emotion ‘iconicity.’ Our analysis will show how the same group of sounds can express different, seemingly incompatible emotional qualities. In another article (Gafni and Tsur, 2019), we present results from a stimulus–response experiment showing that phonemes and phoneme groups do have features with a potential to generate a wide range of sometimes conflicting perceptual qualities that may resemble similar features in a range of emotions. In both articles we attempt to explain why such relations are hard to establish in stimulus–response tests.

A structural theory of phonetic symbolism

The premise of the current proposal is that speech sounds and emotional qualities can be described in terms of more abstract features, and that phonetic symbolism is an association between sounds and emotions through analogy. However, such analogies cannot be straightforward, one-to-one similarities, since there is only a small number of phoneme classes in any language, but an indefinite number of unforeseeable effects: referential meanings, percepts, attitudes and emotional qualities that can be associated with the sound structure of a poem. Therefore, we will argue that if people are looking for “an iconic relation between phoneme classes and emotional classification” (Kraxenberger & Menninghaus, 2016, p. 7), they are bound to fail to find “method in this madness.” Instead, we must look for a more flexible system that relies on the creativity and unforeseeable combinations of symbolic language that is, nevertheless, constrained (rather than determined) by the structure of phonemes. To put it simply, we should look for constraints within which creative mechanisms generate an indefinite number of poetic qualities. This would also account, in a principled manner, for instances when competent readers detect different emotional qualities in the same text, or rigorous statistical studies arrive at incompatible findings. In other words,
in accounting for incompatible effects, we should look for general potentials, rather than arbitrate between specific judgments. ‘Potentials’ suggests having latent (sometimes conflicting) qualities or abilities that may or may not be realized in future contexts. The same sound structures may have conflicting potentials, leaving room for various respondents or studies to realize one or the other of the conflicting potentials, or both.

**Organization of the mental database**

In this section we briefly outline the organization of the mental database indicating the relevant kinds of creativity it affords. Our point is that phonetic symbolism relies on the same kinds of mechanisms as other aspects of linguistic competence.

We have said that in dealing with sound symbolism in poetry we must look for a flexible system that relies on the creativity and unforeseeable combinations of symbolic language that is, nevertheless, *constrained* (not determined) by the structure of speech sounds. This flexibility of symbolic language is due to the fact that the relevant information is not just there in our mental database, but is organized in a certain way and functions in a certain way. Such organization is crucial for the understanding of how linguistic creativity works in general, and phonetic symbolism in particular. We shall focus on two aspects of this organization that are relevant to our inquiry. In short, concepts are bundles of meaning components, or semantic features, or semantic primitives, by virtue of which they are organized into hierarchies called hyponymy. Similarly, phonemes are bundles of distinctive features by virtue of which they are organized into groups and hierarchies of groups. In what follows, we will elaborate on the structures of phonemes and concepts and show how symbolic associations can be formed between them.

---

2 The hierarchic conception of our conceptual system goes back to Aristotle; in modern psychology it is central to Rosch et al.’s work (e.g. Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) on categorization.
How do such bundles of features interact? The creative potential of the aforesaid organization is realized through such linguistic mechanisms as feature abstraction, feature transfer and feature deletion that underlie our linguistic competence. Feature transfer is a regular part of our linguistic competence. When I say ‘My teacher is a bachelor,’ I state that he is unmarried, and specify ‘my teacher’ as a male (i.e. I transfer the feature $<+\text{MALE}>$ to the features of ‘teacher’). When I say ‘My teacher is not a bachelor,’ I negate only that he is unmarried, not that he is a male. The difference between such an exercise and sound symbolism is that, in the latter, semantic features are transferred not into a bundle of semantic features, but into a bundle of auditory features, or an abstraction from them. The other two mechanisms will be exemplified later.

**The structure of emotions**

Some theories of emotion concern the structures and the mechanisms that produce a variety of emotions, some the resulting taxonomies. There has been extensive investigation of the structure of emotions and emotion terms. For our purpose, according to Arnold and Gasson (1968), emotions consist of cognitive situation appraisal plus undifferentiated energy, involving a tendency toward an object judged suitable, or away from an object judged unsuitable. Thus, for instance, by injecting to participants epinephrine, that increased their psychic energy, Lazarus (1968) succeeded to induce in them such disparate emotions as anger and euphoria, by manipulating situation appraisal. Emotions are active in the background, without pre-empting everything else (D’Andrade, 1981). Accordingly, the effect of sound–emotion symbolism is determined by an interaction between ‘cognitive situation appraisal’ expressed by the meaning, a psychological atmosphere abstracted from the sound structure active in the background, and undifferentiated energy. According to gestalt theory,

3 The notion ‘transfer features’ was introduced to linguistics by Weinreich (1966).
emotions are typically related to weak gestalts (Arnheim, 1967; Meyer, 1956); in the
Rorschach test, they are related to such gestalt-free qualities as colour and shading (Alcock,
1963; Rorschach, 1951). In divergent rhythms in poetry, diffuse alliteration patterns, as well
as conflicting patterns of stress and versification that blur each other, generate weak gestalts
and are typically related to emotional qualities.

As to emotional terms, within the *circumplex model* of emotions (Russell, 1980), for
example, emotions can be organized in a two-dimensional space whose axes are *valence
(positive–negative, or pleasure–displeasure)* and *degree-of-arousal* (or *activity*). We
conceive of these two terms as generative principles that can generate an open set of
emotions (as opposed to the closed set of the phonological system of a language).

Other theories (e.g., Johnson-Laird & Oatley, 1989) propose that emotion terms can be
classified into basic emotions, which have no internal semantic structure (e.g., *happiness,
sadness, fear*), and non-basic emotions that are hyponyms of basic emotions (e.g.,
*embarrassment, pity*). According to the theory proposed here, from quite different basic-level
emotions, some general psychological atmosphere may be abstracted or, conversely, such a
general psychological atmosphere may be individuated in (sometimes incompatible) specific
emotions. In phonetic symbolism, similar psychological atmospheres may be abstracted from
sound structures and emotions. Feature analysis is applicable to emotions just as to concepts.

Theories of the semantic structure of emotions can account for what may seem like
conflicting findings among studies of phonetic symbolism in poetry. For instance, in the case
of Fónagy’s (1961) and Wiseman and van Peer’s (2002) categories, ‘grief,’ ‘love,’ ‘pity’ and
many more specific emotions are hyponyms of ‘tender emotion.’ Thus, finding that high
frequency of nasals can be correlated with both a tender and a grieving poetic tone is not
contradictory, since ‘grief’ is a specific form of general tender emotions. As we shall see,
from joy and aggression, too, similar features can be abstracted, that generate similar general
psychological atmospheres (which, in the absence of a better term, we shall call ‘arousal’),
that would account for both being correlated with plosives.

The structure of phonemes

The phonemes of a given language can be classified according to several distinctive
features. For consonants, the relevant features are voice (voiced, voiceless), place-of-
articulation (e.g., bilabial, alveolar, velar), and several other features representing the
manner-of-articulation. One of the relevant manner features is \([±\text{continuity}]\).

Consonants can be abrupt (–continuant): plosives (p, t, k, b, d, g) or affricates (as ts [in
tsar], dž [in John or George], or pf [in German pfuj]); or, they may be continuous
(+continuant), as nasals (m, n), liquids (l, r), glides (w [as in wield], or y [as in yield]); or
fricatives (f, v, s, š [as in shield]). Continuous sounds may be periodic (nasals, liquids, glides)
or aperiodic (fricatives). In periodic sounds, the same wave form is repeated indefinitely,
while aperiodic sounds consist of streams of irregular sound stimuli. Consonants can be
unvoiced or voiced. All nasals, liquids and glides are voiced by default. Voiced plosives,
fricatives and affricates (e.g., b, v, ž, and dž) consist, acoustically, of their unvoiced
counterpart plus a stream of periodic voicing. These are objective descriptions of the
consonants; all language-users have strong intuitions about them, but frequently they cannot
put their finger on precisely what is the object of their intuition.

Another crucial issue about the structure of speech sounds is their relative encodedness:
We consciously hear the unitary speech categories, rarely some of the acoustic information
that carries them. This is called ‘categorical perception,’ or the ‘speech mode.’ If you ask
about the syllables ba, da, ga uttered on the same pitch which one is acoustically higher,
people may answer that they don’t know what you are talking about; only very few people
will be able to tell that the three syllables are arranged in an ascending order of pitch (the
only difference between them being the onset frequency of the second formant). On the other hand, if you ask which speech sound is higher [s] or [ʂ], most people will be able to tell that the former is acoustically higher. Plosives are said to be thoroughly encoded, that is, no precategorical acoustic information reaches awareness (hence compact, tight, and solid in perception); periodic speech sounds and the fricatives [s] and [ʂ] are less thoroughly encoded: In certain circumstances, some of the precategorical sound information may reach awareness. The results of numberless experiments suggest that the precategorical auditory information in lowly encoded phonemes reverberates in short-term memory; that’s why they are less compact, tight, and solid in perception. We should, perhaps, improve on this received view by connecting it with another received view: “In /k, g/ spectral energy is concentrated, whereas in /t, d/, and /p, b/ is spread, with an emphasis on lower frequencies in /p, b/ and on higher frequencies in /t, d/” (Jakobson & Waugh, 1979, p. 103). Velar stops “display a stronger concentration of explosion” than labials and dentals (Jakowbson & Waugh, 1979, pp. 100-101). In our empirical study we found that bilabial consonants were perceived as being ‘softer’ and ‘smoother’ than velar consonants. Bilabial consonants were also perceived as being ‘softer’ than alveolar consonants (Gafni & Tsur, 2019). Perhaps, some of the relative concentration of the plosives is nevertheless perceived subliminally. All vowels are periodic and relatively unencoded. When we have the intuition that [i] is higher, brighter, sharper and smaller than [u], it is because some of the precategorical sound information does reach awareness from behind the unitary speech category. This is the ‘poetic mode of speech perception.’

The structure of speech sounds determines their expressive potentials. Just a few hints: Periodic consonants are more like music, aperiodic consonants more like inarticulate noise; this difference affects their expressive potentials, smooth or rough to some extent or other. This can explain why the aforementioned studies found that ‘tender emotions’ or their
hyponyms are best expressed by periodic consonants. In gestaltist terms, plosives have sharply defined boundaries, and are perceived as less penetrable and ‘harder’ than continuants. At the same time, they consist of energy transmitted outward as a shockwave, and so forth. Thus, they have various perceptual potentials: In changing contexts, listeners may switch between passive hardness and active shockwaves. Voiced stops are ambiguous: They consist of an abrupt plosive plus voicing, which is periodic. In the perceived quality of voiced plosives, the voicing stream may add resonance to the plosive element, or a more massive presence. Thus, in empirical tests, participants may give at least two systematically different responses to plosives in general, and additional two systematically different responses to voiced plosives.

**Principles of sound symbolism**

A non-arbitrary, iconic relation between sound and meaning can be formed when structural similarities are detected between the sound and the intended concept. We claim that speech sounds cannot be ‘iconic’ of specific meanings. First, potentials of meaning and perceptual qualities of the speech sounds are based on opposing perceptual qualities of contrasting phoneme groups that are built into the physical properties of the precategorical auditory information. Second, these potentials suggest not specific meanings, but the highest possible abstractions which, in turn, may be individuated by semantics into lower, more concrete perceptual qualities or meanings. We will elaborate on these two principles in the following passages.

First, we claim, contra Kraxenberger and Menninghaus (2016), that it is the articulatory or precategorical, rather than phonological/categorical, information that is responsible for sound symbolism. Consider Hans Heinrich Meier’s (1999) observation: “while verbs like *to bark* and *to swallow* induce ideophonic interpretation, the homophonic nouns meaning ‘rind of
tree’ and ‘bird of the genus Hirundo’ do not” (p. 138). We are not interested whether the etymology of the verb swallow justifies an iconic construal. Our approach is post hoc, and our point of departure is that some people, including scholars, experience the two homophones differently. In the verb ‘swallow,’ the meaning foregrounds the closing inward movement of the lips in the articulation of the two tokens of w, and the backward movement indicated by the ‘dark’ back vowels; the periodic liquid and glide foreground the smoothness of the process, while the roundedness of the w and of the back vowels4 foregrounds aspects of the movements of the lips and the throat in swallowing. We have enumerated separate features that are experienced as one integrated process. In the bird sense of the word, meaning activates no such iconic relationships. In this way, speech sounds may convey perceptual or emotional qualities by articulatory gestures and precategorical auditory information, beyond their arbitrary referential function. Consequently, in its bird sense, the word is experienced as somehow slimmer. Briefly, the phonological categories act in exactly the same way in the two homophones, but in the verb, features of the meaning activate some perceptual potentials of the precategorical auditory stream and of the articulatory gestures.

Moreover, speech sounds can be used in non-linguistic vocal gestures to convey an attitude. For instance, when mothers hush their babies, they would sound a prolonged [§], a consonant in which rich inarticulate precategorical auditory information is available.

Second, as we have seen, the iconic functions of plosives, for instance, have been extensively explored by statistical means. But always new, unexpected cases may crop up. Consider the following theatre-director’s comments on a single verse line in his Preface to Hamlet. He mentions phonetic elements only briefly, to suggest wider dramatic effects.

---

4 Knoeferle et al. (2017) found that judgments of visual shape [i.e. round-angular] were predicted by vowel formants F2 and F3, which correlate with vowel rounding.
And, while the words themselves say little, the strict scansion, that firm, five times repeated “d,” and the bitten final “n” of

Indeed, my lord, it followed hard upon.

bespeak a Horatio sternly if discreetly of Hamlet’s mind about it. (Granville-Barker, 1957, p. 210)

Granville-Barker (1957) does not describe Horatio’s speech as aggressive, as we would expect based on Fónagy’s (1961) findings, nor as joyful, as we would expect based on Wiseman and van Peer’s (2002) and Auracher et al.’s (2011) findings; Granville-Barker describes his direct impression, which is quite different. As we said, sounds cannot be icons of meanings or of sophisticated qualities; they can only suggest elementary sensory or perceptual qualities. We may point out such acoustic properties of /d/ as ‘abrupt’ and such perceptual qualities as ‘hard.’ The rest involves transformations done by the reader, within the constraints of the text. In this case, five /d/ s insert abrupt discontinuities into the text. This renders the text disjointed, not flowing smoothly. On the syntactic level, this is reinforced by the commas separating the words ‘indeed,’ and ‘my lord’ from the main sentence. All this may be INDIVIDUATED, by feature-addition, as ‘careful in speech,’ which is one of the dictionary meanings of ‘discreet.’ From the point of view of contents, “the words themselves say little” suggests that the discreet Horatio is trying to compromise himself as little as possible by nevertheless telling his mind. ‘Strict scansion’ is usually perceived as witty, or simplifying, or hypnotic, but here it rather may reinforce the impression of careful speech. Thus, the perceived quality is generated by exploiting converging potentials of speech sounds, meaning, syntax and rhythm. It is not the effect regularly attributed to these structures, but is overdetermined at various levels of organization. Wiseman and van Peer (2002) and Auracher et al. (2011) think that they are right (and Fónagy [1961] probably wrong) in relating plosives to joy rather than aggression. But plosives can be related only to
such elementary percepts as hardness, abruptness and sudden outward pressure. All the rest is done by similar transformations.

The Case of Onomatopoeia

It has been observed about the foregoing argument that

The writers say several times that sounds cannot be icons of specific meanings. But what about classic onomatopoeia, which even Saussure mentions (though begrudgingly)? Would the writers agree that onomatopoeia is a case where sound is an icon of a specific meaning, namely the meaning of the extra-linguistic referent?

Using a kitchen language, we would agree. Using an analytic language, the process is far more complex. To put it simply, in onomatopoeia, as in any other phonetic symbolism, speech sounds may display some elementary percepts that can be exploited for reference. Onomatopoeia is a special case of sound symbolism because it is an attempt to mimic actual sounds in the extra-linguistic world. If we use the speech sounds to denote some elementary percept, there may be complete identity between sound information and meaning; otherwise there is only a partial identity that is treated as a complete identity. We learn words as a package (don’t infer meaning from sound; at best, sound facilitates the learning of the package). Speech sounds have a wide range of perceptual potentials; the meaning selects the relevant potentials. When the same speech sounds or speech sounds belonging to the same groups are repeated, a general atmosphere may be abstracted from them which, in turn, is individuated into specific emotions or attitudes or percepts or referential meanings by semantic feature transfer from the meaning. In most instances, this feature-laden general atmosphere lingers in the background like pervasive emotions, without pre-empting
everything else. In the case of onomatopoeia, the output of this process is hypothesized to have fossilized in a compact dictionary entry. This would be in harmony with Tsur’s thesis in his recent Oxford UP book, _Poetic Conventions as Cognitive Fossils_ (Tsur, 2017b). De Saussure may now rest in peace. According to the present conception, in symbolic language signifiers are arbitrary; sound symbolism is typically generated _after the event_, as a result of the foregoing cognitive process, based on the _partial_ identity of auditory and semantic features. In some instances, this symbolism fossilizes to some extent or other. In genuine onomatopoeia, identity of features and fossilization are more complete, a kind of ‘frozen metaphor.’

Couplings of speech sounds and meaning didn’t descend to us from Mount Sinai like the Torah, ready-made. Imagine one of our remote ancestors making a contribution to the alleged protolanguage said to be essentially onomatopoetic: when inventing an onomatopoetic word, that ancestor turns a partial identity into a complete identity. If s/he coined the verb ‘murmur’, s/he did this because s/he heard some inarticulate precategorical low-pitched and low-energy sound, continuously repeating, behind the abstract speech categories. If s/he applied it in the sense of ‘low, inarticulate noise,’ the referent _was_ an elementary percept, and there is an almost complete identity between the acoustic and semantic features. This is the most complete identity that can be achieved. But even then, a stream of lowly differentiated periodic sounds refers to a stream of lowly differentiated irregular noise. At this point, “standard semantic extension (such as generalization, specialization, metaphor and metonymy) can account at least partially for some degree of semantic shift,” as has been said in a slightly different context (Smith, 2016, p. 8).

If that ancestor applied it in the sense of ‘a low continuous background noise,’ however similar, the meaning contained features that the speech sounds didn’t. If many thousand years later, descendants used the word in the sense of ‘a softly spoken or almost inaudible
utterance,’ it is still onomatopoeia, but many more features must be added to the perceptual features abstracted from the speech sounds. When ur contemporary readers read Pope’s “If Chrystal Streams with pleasing Murmurs creep,” they apply the latter meaning, by deleting the semantic feature [+HUMAN], obtaining a metaphor for the sound of the stream; or transfer the feature <+HUMAN> too to the ‘Chrystal streams,’ obtaining personification.

Accommodating conflicting findings

We have mentioned the inconsistent results of research done by Auracher et al. (2011), Fónagy (1961), and Wiseman and van Peer (2002). Iván Fónagy (1961) found that the unvoiced stops /k/ and /t/ are significantly less frequent in tender poems than in aggressive poems by Petőfi, Verlaine, Hugo, and Rückert (Hungarian, French, and German poets). Wiseman and van Peer (2002) found that “nasals and back vowels seem to be preferred to express grief, while joy is better expressed by plosives and front vowels” (p. 382). Although they followed different methodologies, this is not responsible for this inconsistency. Even though both were looking for correlations between phoneme groups and pairs of opposite emotions, they investigated different pairs of opposing emotions; and, not surprisingly, found correlations between the same phoneme groups on the one hand, and different pairs of emotional qualities on the other. Tsur (2012) explored whether the two can be reconciled (pp. 271–274). Suppose Wiseman and van Peer (2002) asked their informants about nasals versus plosives, inquiring whether they express tender or aggressive emotions; they would have got the answer that nasals are tender and plosives aggressive. How do we know this?: because such results are supported by Fónagy’s (1961) explorations. One finding does not prevent the other from being accurate. Now, it is easy to see why nasals can express both grief and tender

---

5 Auracher et al. (2011) hoped to establish that one can predict the emotional quality of a poem from its phoneme structure; whereas Fónagy explored the opposite direction.
moods, because grief is a tender emotion. It is more difficult to see how plosives can express both joy and anger. This you can explain only by analysing the phonetic structure of the speech sounds and the structure of the emotions.

We conceive of phonetic symbolism as a function that cannot indicate specific emotions, only very general emotional atmospheres which, in turn, may be individuated into specific emotions by semantic features. As a matter of fact, we are faced with a dichotomous spectrum of very wide categories, one pole of which is marked as low-energy, inward-turning emotions, characterized by lack of sudden sharp changes; the other pole is marked by high-energy, outward-turning emotions, sometimes characterized by outbursts of varying degrees. At the former end, we locate emotions such as ‘tenderness,’ ‘grief,’ ‘love,’ ‘pity,’ and others. Such emotions typically bear a structural resemblance to glides, liquids, and nasals.

At the other end of the spectrum, we have an inclusive category that contains both ‘joy’ and ‘anger’ which, for want of a better name we tagged ‘arousal,’ even though they share features to some of which the tag is irrelevant; they may share high energy levels, outward turning and outbursts, that is, energy transmitted outward like a shockwave — not unlike plosives in phonetics. Energy can be individuated by semantics in anger as destructive, in joy as exuberant. The abruptness of plosives suggests clearly articulated body language in joy, whereas in aggression it may suggest somewhat stiff (that is, over-articulated) body-language, not flowing smoothly. Likewise, the smoothness and softness of [l, m, n] can be tainted in grief by a sense of withdrawal, whereas in love and pity, with a sense of drawing toward.

Our analysis can also explain one of Fónagy’s (1961) most intriguing findings regarding the relative frequency of /g/ and /d/ in Victor Hugo’s and Paul Verlaine’s poems. /g/ occurs

---

6 There is a theatrical tradition of portraying the threatening alazon with stiff movements; this is a live attitude fossilized into style.
over one and a half times more frequently in Verlaine’s tender poems than in his angry ones (1.63:1.07), whereas we find almost the reverse proportion in Hugo’s poems (0.96:1.35). As to /d/, again, the same sound has opposite emotional tendencies for the two poets, but with reverse effects. For Verlaine it has a basically aggressive quality (7.93:10.11), whereas for Hugo it has a basically tender quality (7.09:5.76) – again, in almost the same reverse proportion.

At first glance, this pattern of results may seem arbitrary. However, we argue that this is not so. The reason seems to be what Wittgenstein (1976) calls ‘aspect-switching’ (p. 214e). If you attend to voicing in /g/ or /d/ as what bestows massive presence on the plosive ingredient, it may have a strong aggressive potential; if you attend to it as what adds resonance to the plosive ingredient, it may contribute to a tender quality; in other words, you may change the relative weight you attribute to the same elements. Wittgenstein applies the term ‘aspect-switching’ to the semantic ingredient of words, as pronouncing the word ‘March’ and understanding it as a month name or an imperative verb; the same switching can be applied, we argue, to the effects of voicing in their phonetic ingredient. Tsur (1987) has elsewhere demonstrated that in Wittgenstein’s visual examples (seeing as), aspect-switching is a matter of feature-addition (pp. 101-103). Obviously, Hugo and Verlaine applied the same cognitive mechanism to these voiced stops, but with a reverse focus.

Thus, we have shown how the same group of sounds can be associated with different and discordant, specific emotions. But our analysis can also accommodate the more general conflicting findings on the valence (or, hedonic) scale, such as those associating plosives with both positive (or, pleasant) emotions (Auracher et al., 2011) and negative (or, unpleasant) emotions (Whissell, 1999). Again, we claim that the conflict is only at the surface. Depending on their internal structure, both positive and negative emotions can have high or low degrees of energy. In the former case, they are more likely to be associated with abrupt
consonants, in the latter with (continuous) periodic consonants. Thus, structural similarities among different emotions can account for their potential to be associated with the same group of sounds. The problem arises when one attempts to apply an evaluative rather than a descriptive classification of emotions. Sounds, unlike emotions, have no valence; they cannot be positive or negative. Therefore, any association of sounds with emotional valence can only be epi-phenomenal to the examined emotions. In other words, all talk about positive and negative emotions associated with speech sounds is deceptive.

**Objection to the present proposal**

Kraxenberger and Menninghaus (2016) comment on our treatment of Fónagy’s (1961) and Wiseman and van Peer’s (2002) contradictory results that “Since this interpretation, however, is not based on empirical data, the reported contradictions remain unresolved” (p. 3). To that, we answer that what Tsur was doing was nothing that could be supported or refuted by some stimulus–response questionnaire. Fónagy (1961), Wiseman and van Peer (2002), and Auracher et al. (2011) provide ample empirical data that nasals can express grief as well as tender moods, and that plosives are adequate for expressing both joy and aggressiveness. The question is not whether the data are empirical enough, but whether both can be right, and on what grounds. No number of subjects can settle this question. To be sure, it is legitimate to demand ascertainment of one’s assumptions about the person-in-the-street’s intuitions regarding the perceived qualities of speech sounds and emotions, but the argument stands in its own right. At any rate, we have now provided such an empirical study too (see Gafni & Tsur, 2019).

Beyond other inaccuracies, Kraxenberger and Menninghaus seem to presume that there is only one way to validate the results of some research: stimulus–response questionnaire. There are many more, of which we shall mention here only two: thought experiments and case
studies. Adapting Kuhn’s (1977) discussion of Galileo’s thought experiment with the concept of speed, one may try to better understand where the discrepant results of rigorous studies went wrong. In Kuhn’s (1977) words, their defects lay “in [their] failure to fit the full fine structure of the world to which [they were] expected to apply. That is why learning to recognize [their] defects was necessarily learning about the world as well as about the concept” (p. 258).

Some researchers are satisfied with any result to which statistics may lead. In our conception, the first step to validate a research study is to test whether its results fit the full fine structure of reality to which it was expected to apply. Only after that, one may or may not examine whether the proper statistical manipulations were performed in the proper order. In the present instance, we had incongruent results from three rigorous statistical studies. Obviously, some or all of them could not fit the full fine structure of the world to which they were expected to apply. Tsur’s analysis rendered them congruent and having a better fit to the full fine structure of the world.

**Non-reductive Statistical Investigation**

So far, we proposed a critical review of several studies of phonetic symbolism in poetry. Common to all of them (except for Fónagy, 1961) is the simplistic expectation that the emotional quality of a poem can be predicted exclusively based on the relative frequencies of various phonological groups. In other words, studies adopting this approach reduce their subject matter to its lowest common denominator. We claim that the conflicting findings among these studies suggest that the simplistic approach is inadequate. Yet, some studies adopt a much more sophisticated statistical approach that serves as a better candidate for a quantitative model of phonetic symbolism in poetry.
Aryani, Kraxenberger, Ullrich, Jacobs, and Conrad (2016) and Ullrich, Aryani, Kraxenberger, Jacobs, and Conrad (2017) studied the relation between general affective meaning and the use of particular phonological segments in 57 poems (“verteidigung der wölfe”) by the German poet Hans Magnus Enzensberger. The poet himself categorized the poems as ‘friendly,’ ‘sad,’ or ‘spiteful’ (thus, the classification of poems by emotional quality did not depend on judgments by naïve subjects or even the researchers). Participants in these studies rated the poems on various scales (valence, arousal, sad, friendly, liking, poeticity, etc.). Then, the authors conducted statistical analyses in attempt to predict the subjective ratings, as well as the poet’s own classifications, based on the properties of the poems. However, unlike previous studies, Aryani et al. (2016) and Ullrich et al. (2017) did not attempt to establish a simple correlation between the perceived quality of a poem and the raw frequencies of phone groups. Instead, their statistical models included a large number of predictors that could contribute to the perceived quality of a poem. For example, they used an external database of German words rated for arousal and valence and calculated the affective potentials of phones based on their distribution in the database (e.g., a phone present in many words that were rated high for arousal would have a high arousal potential). Thus, both studies produced sophisticated statistical models, which demonstrated the relative contribution of multiple factors to the perceived quality of poems. Across both studies, multiple predictors reflecting the sound contents of the poems were statistically significant. Thus, multi-predictor statistical models are far more adequate for investigating phonetic symbolism in large-scale databases.

While we are sympathetic with this approach, we took a different course in an attempt to avoid reductionism, focusing on the multivalence and creative manipulation of features — semantic, phonetic, visual, etc. In short, we take a localist, “micro” approach to phonetic symbolism in poetry, as opposed to the globalist, “macro” approach taken by Aryani et al.
(2016), Ullrich et al. (2017), and others. We have adopted a thorough interaction approach that attempts to provide a cognitive account for phonetic symbolism in poetry. Writing and reading poetry are problem-solving activities. In an attempt to solve the arising problems, we have analysed the semantic, phonetic and versification components of our texts, as well as their emotional qualities, tracing their interaction at a fine-grained level. By this, we claim, we have tapped some of the subtleties of the texts. By contrast, studies following a “macro” approach analyse poems as wholes and, while they are able to point to major factors contributing to the perceived qualities of poems, even the most sophisticated statistical analyses miss some of the subtleties of poetry. Poetic effects result from processes of creativity, in which semantic and phonological components are abstracted, contrasted, and recombined in unforeseeable ways. A rigorous statistical analysis over a large number of poems can only tell us whether there are some predominant factors that are correlated with the emotional quality of poems. It cannot (always) tell us why a given poem is perceived the way it is, which is the topic of many literary studies.

Case studies

The remedy seems to be, then, thought experiments and case studies. As an example of a case study, let us discuss two versions of the following stanza from Fitzgerald’s “The Rubáiyát” of Omar Khayyám (Fitzgerald, 2008):

[3] Some for the Glories of this World; and some
Sigh for the Prophet’s Paradise to come;
Ah, take the Cash and let the Credit go,
Nor heed the rumble of a distant Drum!

(Fifth edition)
[4] “How sweet is mortal Sovranty!”—think some:
Others— “How blest the Paradise to come!”

Ah, take the Cash in hand and waive the Rest;

Oh, the brave Music of a distant Drum!

(First edition)

We are discussing the fifth edition, only comparing it to the first one; it is more complex and more heavily laden with repeated sound patterns. Consider the three rhyme words, some–come–Drum in this version. Some readers report that they are aware of a rich body of reverberating auditory information in Drum, but are not aware of a similar richness in the preceding rhyme-fellows. To be sure, traditional criticism has an excellent explanation for this, as far as it goes. There is an exceptionally rich alliteration pattern in heed–rumble–distant–Drum. The phrase the rumble refers to continuous deep resonant sound; the repeated phonemes /r/, /ʌ/ and /m/, in turn, are perceived as somehow imitating this sound. The present discussion purports to go two steps further, and invoke the rich precategorical auditory information on the one hand, and the fusion of such auditory information on the other.

There is an interesting phenomenon in sound–meaning relationship that has not yet been sufficiently explored. There is disagreement in research whether the phonological signifier of meanings is arbitrary. There is a tradition going back to Ferdinand de Saussure claiming that the phonetic sign is arbitrary; and a rival tradition, going back as far as Plato’s “Cratylus,” claiming that the existence of phonetic iconicity challenges the Saussurean conception (see, e.g., Whissell, 2004). We argue that things are not as simple as that. There are instances in which exactly the same sequences of speech sounds in homonyms or near-homonyms are arbitrary signs or ‘iconic.’ What is more, the latter are not merely iconic, but sound
differently, they sound as having a reverberating, fuller body, even if pronounced in the same way.

We have discussed above at some length the word *swallow*. As a verb, we found an intense interaction between its sound and meaning, adding a fuller body to the sound structure; in its homonym, the bird’s name ‘swallow,’ the same sounds remain flat. In ‘rumbling of a distant drum,’ both the liquids and nasals and even the voiced plosives are perceived as reverberating, owing to the interaction between the sounds and meaning.

It is interesting to compare the last line of the two versions of the Rubáiyát: The repeated sound patterns generate onomatopoeia, foregrounding different aspects of the meaning. In Excerpt [4], the repeated abrupt /d/ foregrounds the focused beat in beating the drum; in Excerpt [3], the repetition of the continuous and periodic /r, m/ reinforced by /l/ foregrounds the reverberating sound of the drum as well. Semantically, line 3 treats, in both versions, the opposition between this and the other world in the down-to-earth terms of money. Phonetically, the alliteration of the abrupt voiceless plosive /k/ in Excerpt [3] foregrounds and sharpens the opposition between immediate and (relinquished) future payment. There is an interesting semantic difference between line 3 and 4. In line 3, the other world is expressed in terms of an intellectually grasped abstract value (credit). In line 4 it is expressed in an immediately experienced undifferentiated sound percept coming from beyond the perceptible horizon. In this respect, there is a significant aesthetic difference between the two versions. In Excerpt [3], the immediate perception of this undifferentiated sound percept is “heightened, to any degree heightened” (to use Hopkins’ words) by the rich precategorical sound information enhanced by the sound repetition in ‘rumble – distant drum.’ In other words, there is here not merely a distinction between the presence and absence of alliteration; the sound repetition ‘rumble – distant drum’ enhances the immediate perception of the undifferentiated sound percept coming from beyond the perceptible “Ultimate Limit.” Still, in
other words, this becomes a verbal imitation of the undifferentiated intuition of the world beyond the Ultimate Limit. In the first edition, only the abrupt /d/, not the reverberating liquid and nasal, is repeated.

There is a thoroughly experimented phenomenon in speech perception, lateral inhibition (Crowder, 1982a, 1982b), indicating that for biologically motivated reasons too, the same speech sounds are sometimes more and sometimes less reverberating. When speech sounds in proximity are very similar, they enhance each other’s precategorical sound information in perception; when they are slightly similar, they inhibit each other; when they are not similar at all, there is no mutual effect. In other words, in some instances alliterating speech sounds may enhance reverberation in each other, in some — inhibit it.

Robert G. Crowder suggests (personal communication) that there would be precedent for the assumption that the total effect would be the larger for having had a repeated sound. This depends on his assumption that both inhibitory and enhancing interaction takes place within the formant energy of the words, even though they may be spoken at different pitches (formants are concentrations of overtones that uniquely determine vowels; cf. below Ehrenzweig’s discussion of overtone fusion in music). Here, in the phrase ‘rumble of a distant Drum’ there are very similar groups of speech sounds, effectively enhancing each other’s formants (overtones).

7 ‘Distant’ places the words ‘rumble...drum’ in a spatial perspective. Words are related to the left hemisphere of the brain, whose output is logical and sequential; whereas emotions and intuitions as well as space perception and sound perception are related to the right hemisphere, whose output is global and diffuse. The spatial perspective amplifies the diffuse emotional and intuitive quality of the inarticulate sounds, suggesting the perception of an invisible and inaccessible reality. That is why romantic and symbolist poets tend to express their emotion or the intuition of an invisible, inaccessible world through landscape descriptions — sometimes without using emotion words at all.

8 One of the chapters in Tsur (1992) is called “Musicality in Verse and Phonological Universals.” It took him years to discover that the sound sequence /vǝrs/ occurs in it twice, a powerful alliteration, as it were. In ordinary speech, the same sound sequence is longer in monosyllables than in polysyllables (as in ‘tail—tailor’ [see Fónagy & Magdics, 1960]). Thus, the second token of /vǝrs/ is much shorter than, less similar to, the first token, exercising lateral inhibition. One must, therefore, lengthen the second token to equalize it with the first one in order to hear the alliteration.
The liquids and nasals /r/, /l/ and /m/ and vowels in general are much less encoded than the voiceless plosives /p/ and /k/; that is, in this phrase some precategorical auditory information is perceptible. Furthermore, the meaning of the phrase foregrounds this precategorical auditory information, generating a very effective onomatopoeia. Moreover, by the same token, it also foregrounds the reverberations of voicing in the voiced plosives /b/ and /d/. In other words, the reverberation of certain acoustic features is enhanced in perception by similar semantic features.

In the earlier rhyme-fellows, ‘some’ and ‘come,’ the auditory information of the /ʌ/ and /m/ is less enhanced (they are further away, and don’t participate in an additional sound pattern) and are not foregrounded by meaning. Moreover, the enjambment in the first two lines of this Rubáiyát (fifth edition) is characterized by what Tsur (1972) called ‘articulateness’ and ‘requiredness,’ based on the Gestalt assumption that a perceptual unit tends “to preserve its integrity by resisting interruptions” (Fodor & Bever, 1965, p. 415). If a major syntactic boundary intrudes upon a verse line in the middle, it enhances its balance; the nearer to the end, the stronger the sense of disruption caused, and the stronger the sense of relief when the missing part clicks into place. Moreover, a push is perceived from the syntactic interruption toward the line ending. This holds true for end-stopped lines; in run-on lines, the intruding event generates an intense forward push to the next line (see Tsur, 1972). In the present instance this weakens the rhyme at the end of the first line, and achieves exceptionally strong focal stability on the second rhyme word, in spite of the continuous, periodic nasal. This also may render the repeated sounds in the rhyme words sufficiently different to induce the two rhyme words to inhibit each other’s reverberation. It may be

---

9 In Fodor and Bever’s experiment, participants listened to a complex sentence, in course of which a click occurred. After listening, they had to indicate where did the click occur. They were most accurate when the click occurred at the clause boundary. If it occurred before or after the boundary, they pushed it in their memory toward the boundary. The purpose of the experiment was to test whether syntactic units and their boundaries have psychological reality. From their findings we may extrapolate to versification boundaries as well.
illuminating to compare the first two lines of the two versions. In the first line of both versions, a major syntactic boundary intrudes upon the verse line after the eighth position, near the end, exerting powerful forward push. In Excerpt [4], this push enhances the closure of the line, sharply separating the two lines. In Excerpt [3], this push enhances both the fluidity of the two lines and the requiredness of the second line, generating a powerful closure. It should be noted that here the boundary interrupts not only the verse line, but the syntactic unit too, enhancing the requiredness of the next line, where the shared predicate of the two sentences occurs; nevertheless, this exceptionally fluid structure is well articulated, owing to the quasi-chiastic repetition of ‘some,’ as the first and last word of the line.

Now notice the alliterations Prophet’s Paradise and Cash–Credit. The voiceless plosives are highly encoded, and no precategorical auditory information can be perceived in them (hence compact, tight, and solid in perception). So, they are perceived as less reverberating, ‘opaquer,’ and having a ‘leaner,’ clear-cut body. They are foregrounded, because each pair occurs in stressed syllables in strong positions and within strong gestalt boundaries (see below) and, as such, are particularly snapping and cracking. However, they do not serve as onomatopoeia, nor do they express joy or anger. In this state of affairs there are two possibilities, both deviating from the dominant view today. Either they represent Hrushovski’s (1980) category of neutral sound patterns, or they are expressive of some unforeseen quality abstracted from the meaning. For instance, the line “Ah, take the Cash and let the Credit go” may display an assertive, conclusive tone that may be reinforced (or expressed) by the sharply defined boundaries of the voiceless plosives. But, it would appear, ‘Prophet’s Paradise’ must be treated as a neutral sound pattern.

According to the Gestalt psychologists, “colour interaction increases within the boundaries of a good Gestalt while it is inhibited across its borders”; the same holds true of overtone interaction in music (Ehrenzweig, 1970, p. 172) and, we claim, of overtone interaction in
poetry. The alliterations ‘Prophet’s Paradise’ and ‘Cash–Credit’ occur in end-stopped lines, within strong gestalt boundaries. Since the alliterating plosives occur in stressed syllables in strong position, opposed to adjacent unstressed syllables in weak positions, they further strengthen the gestalt. Furthermore, since the two alliterating words occur on the two sides of the caesura, they enhance symmetry, further strengthening the gestalt. Since, in voiceless plosives, no precategorical auditory information can be perceived, no overtone interaction can take place; it is rather their sharply defined boundaries that collide, amplifying their snapping-cracking sound pattern.

Auracher et al. (2011), however, relying on statistical correlations, not structural resemblance, claim that \( P \) stands for happiness (that could be quite suggestive in the context of ‘Prophet’s Paradise’); so, if it is happiness, here it is rather snapping happiness. To put it more explicitly, if \( P \) does, indeed, stand for happiness in sound symbolism, it can only be due to the relative softness (relative diffuseness of its acoustic energy) as compared to \([t, k]\). The gestalt dynamics described in the preceding paragraph, however, imposes focused perception on this diffuse energy, rendering it brisk and sharp. This alliteration bestows a sense of sharp boundaries, epigrammatic certainty on the verse line, rather than a sense of fuzzy or suspended boundaries characteristic of happiness.

In the last line, on the contrary, the rich auditory information intensely interacts in the alliterating words and seems to soften the overall impression, amplifying its full resounding quality. As Ehrenzweig (1965) said, “As in all relationships between form and colour the reverse effect can also happen. Strong colour interaction [in this case, overtone interaction] tends to make sharp outlines seem much softer than they are” (pp. 172-173).

A wide range of theoretical distinctions expounded throughout our discussion of phonetic symbolism are focused on this example. We assigned structural descriptions to sound repetitions in this poem and made claims about plausible intuitions regarding their effects,
without adducing any evidence that anybody ever responded to them in this way. We supported our claims by hypotheses based on empirical research on relevant perceptual processes in gestalt psychology and speech research. Thus, for instance, we claimed, based on gestalt theory and our own introspection, that a syntactic boundary intruding near the line boundary exerts a ‘perceptual force’ pushing forward toward the line boundary. We support our claim by Arnheim’s (1967) demonstration of the phenomenon in visual perception (p. 2), and by Fodor et al.’s (1965) click experiments, which were originally meant to support the psychological reality of syntactic units (including boundaries).

What we have pointed out in the sound repetitions of this quatrain is merely a potential. ‘Empirical’ studies, on the other hand, merely determine how many respondents do actualize this potential or part of it. Their results will be crucially influenced by the structure of the text, the respondents’ personality style and theoretical sophistication, as well as the experimental design. So, if they don’t realize those potentials of the poem, it does not necessarily give information about the poem, but about the respondents’ poetic competence. One of the roles of critical communication is to draw attention to those potentials rather than to prove anything. The role of theory in aesthetics is, in Morris Weitz’s (1956) terms, “not to define anything but to use the definitional form, almost epigrammatically, to pin-point a crucial recommendation” as to what to look for and how to look at it in art (p. 35).

Thus, we made a crucial recommendation to notice some perceptual differences in the rhyme words ‘some–come–drum’ and some perceptual differences between the alliteration patterns ‘Prophet’s Paradise,’ ‘Cash… Credit’ on the one hand, and ‘rumble… drum’ on the other. We also pointed out the phonetic structures of those sound patterns, their interaction (or lack of interaction) with meaning, and the processes of speech perception and gestalt dynamics involved (as established by non-aesthetic empirical research). Finally, as an act of (plausible) interpretation, we pointed out that ‘the rumble of a distant drum’ may serve as an
exceptionally vivid sensuous metaphor for the (intuitive) perception of (undifferentiated) information coming from beyond the Ultimate Limit. As Tsur (2003, 2018) pointed out, in some religious poetry, the otherworldly or the numinous is perceived through undifferentiated percepts rather than stable visual shapes. In other words, rather than testing whether people noticed those subtleties of the text, the case study may help them, in the first place, to realize them.

**Theory adequacy and empirical results**

Earlier, we showed how the same group of phones can contribute to the expression of different emotions, and thus accommodate some conflicting findings from previous studies. It is less obvious how the null finding of Kraxenberger and Menninghaus (2016) can be accounted for. Kraxenberger and Menninghaus examined the correlations between the sound contents and emotional qualities of 48 German poems, written by various authors over the course of 150 years (mostly in the 20th century). They did not find significant correlations between the emotional quality of the poem (sad or joyful) and the frequencies of nasals or plosives. This stands in sharp contrast to most studies on the topic. So, how can the findings of Kraxenberger and Menninghaus be accounted for in light of previous studies?

Kraxenberger and Menninghaus performed elaborate statistical analysis on a large, relatively diverse database. A priori, we would expect their analysis to find a significant effect, if such exists.

In one of his Minnesota seminars in the 1970s, Walter Weimer claimed that Sir Karl Popper’s *The Logic of Scientific Discovery* is a misnomer. The book ought to be called ‘The logic of the completed scientific report.’ Scientific discoveries follow a different logic. In Kraxenberger and Menninghaus’ (2016) published article we face the logic of the completed scientific report. We may hope to find traces of the logic of *discovery* in the first author’s
PhD dissertation, on which the published article is based. We have usually no documentation of the researchers’ intuitive decisions in constructing their research plans. But in the present instance, it seems to us, we have a tangible Ariadne’s spool leading to this mystery.

Kraxenberger (2017) sums up her relevant findings in her dissertation as follows:

[T]he results of this dissertation lend support to the idea of sound–emotion associations in poetry. In contrast to the relation posited by previous investigations, in this project such a relationship does not seem to be dependent on the frequencies of occurrence of certain phonemes. However, this dissertation shows that figures of phonological recurrence (alliteration, assonance, consonance) are perceived as distinctively joyful. […] Based on the empirical results of the dissertation, it is argued that sound–emotion associations in poetry are primarily semantically motivated, i.e., to depend substantially on the linguistic competence of the reader to comprehend and access a poem’s semantics. (p. 154)

So, how can it be that “figures of phonological recurrence […] are perceived as distinctively joyful” but not distinctively sad? What is the reason of this asymmetry? We believe that the problem is built into the texts, but does not show up in an inadequate theoretical approach. Kraxenberger does not distinguish between stress pattern and versification pattern. As a matter of fact, she does not speak of stress pattern or versification pattern at all, only marks word stress. She does not treat linguistic stress or accent as part of poetic rhythm, only as something that may or may not coincide with participants’ emotionality judgments: as if poetic rhythm had nothing to do with the poem’s emotional quality. In the last paragraphs of her conclusions section, though, she acknowledges the importance of poetic rhythm to emotion, subject to future research.

In many of his writings, since 1972, Tsur (1972; 2008, pp. 59-61, pp. 83-91, pp. 100-104; 2017a, pp. 16-19) has distinguished between convergent and divergent rhythms. Convergent
structures have strong gestalts and are intimately related with witty poetry, simplified mastery of reality (as in nursery rhymes and folk songs), and with hypnotic poetry. Divergent structures have weak gestalt, and are typically related with emotional qualities like sadness. We claim that this distinction takes care of the rhythm’s effect on the emotional qualities of poetry. (At any rate, it would appear that in German it is more difficult to find divergent poems than in English.) There is much evidence that even intensive sound patterns are less easily discernible in divergent structures than in convergent structures. Consider, for example, the second quatrain in Baudelaire’s “Correspondances”:

Comme de longs échos qui de loin se confondent
Dans une ténébreuse et profonde unité,
Vaste comme la nuit et comme la clarté,
Les parfums, les couleurs et les sons se répondent.

Like long echoes that mingle in the distance
In a dark and profound unity,
Vast as the night and as the light,
The perfumes, the colors, and the sounds respond to one another.

Henri Peyre, professor of French at Yale and author of a monograph on Baudelaire (excellent credentials for being a competent reader of Baudelaire), wrote in his comments on Baudelaire’s sonnet “Correspondances”: “There is no virtuoso musical effect in the quatrains” (Burnshaw, 1964, p. 9). Tsur (2008), however, pointed out an exceptionally rich and intricate but diffuse sound texture in the second quatrain (p. 497):

In line 1, the word longs is entirely contained in loin, while it forms a ‘sporadic rhyme’ with sons in line 4; it also has the nasal vowel -on in common with confondent ... profonde ...
répondent. The first and third of these three words make a formal rhyme as expected, while the second appears unexpectedly in midline, what Roi Tartakovksy (2014) calls ‘sporadic rhyme.’ The first three consonants of profonde reappear in parfums in the same order; the first two are found (in a reversed order) in répondent, the third in confondent. To this, one might add Dans une ténébreuse et profonde unité, remembering that /b/ and /p/ differ only in voice. The sounds of unité are anticipated by une and ténébreuse, while its first three sounds are re-sounded in nuit (in a different order) in the next line. Clarté, while rhyming with unité, contains all the consonants of couleurs, in the same order. Most of these repetitions are sound clusters, involving two or three, elements. The clusters are perceived and subliminally compared to each other; the reader, however, cannot focus awareness on any particular string, because attentive perception has been distracted from one string to another, so that a network of highly significant sounds has been generated, rich in effects, but only semi-consciously perceived. (For another outstanding example, in English poetry, see Tsur, 2017, pp. 17–19).

We don’t know to what extent is this phenomenon of backgrounded, diffuse but intense sound repetition widespread in German poetry; at any rate, in Milton and in Baudelaire, it is very effective. If in German poetry they are non-existent, we cannot infer from German to French and English poetry. Even if they exist, in view of Kraxenberger’s (2017) explicit criteria for marking her categories, most of these sound patterns would slip through her fishing net. Peyre seems to have missed these sound patterns because, in spite of being exceptionally intense, they are also exceptionally divergent (with naïve subjects we may have the additional problem of lacking the required poetic competence) (Burnshaw, 1964). We cannot know whether other researchers intuitively make the convergent–divergent distinction, but it is a good candidate for accounting for Kraxenberger’s puzzling results.

It could be objected that other researchers don’t use either the convergent/divergent distinction in this context. Why, then, other researchers should reach different conclusions?
Our point is that Kraxenberger and Menninghaus’s published article (2016) gives the logic of the completed scientific report, from which we can learn nothing about their way of thinking. The dissertation (Kraxenberger, 2017) gives us some clue how their methodology precluded significant insights into the nature of the relation between speech sounds and emotions. The only poem in which Kraxenberger (2017) shows, in two lines, how she handled sound patterns is exceptionally convergent and judged joyful. She puts assonances, consonances and alliterations in different rows of her table, so that you cannot know whether there is any structural relationship between them, certainly not whether they converge or diverge. The only thing you can do with them is statistics. On the surface, Aryani, Conrad, Schmidtke, and Jacobs (2018) follow the same logic of the completed scientific report. But the subtleties they found give evidence that they were sensitive to the interaction of phonetic and semantic features, though they don’t say so. They do, indeed, make not only such distinctions as ±obscene, but also such as *piss* is more obscene than *pee*. The former distinction can be made by considering semantic features alone; the latter only by considering the interaction between semantic and phonetic features. They’ve also found that the same voiceless plosives and sibilants suggest terror in other poems. But they don’t relate between the two examples. They don’t raise issues related to convergence/divergence, so they need not make this distinction. But Kraxenberger (2017) found that sound patterns are effective in joyful (typically convergent) poems, but not in sad (typically divergent) poems. This curious finding cannot be accounted for without understanding the nature of convergence and divergence.

Before proceeding, we must briefly discuss another issue. Our inquiry concerns the aesthetic domain, not the social sciences. An aesthetic object can be viewed in two perspectives: as a bundle of conventions shared with other aesthetic objects or as the unique combination of conventions that determine its aesthetic uniqueness. Large-scale statistical investigations in poetry indicate, at best, general tendencies, but are perfectly helpless in front
of the aesthetic uniqueness of a poem. Thus, for instance, there may be a poem in which statistical investigation does not turn up any significant sound patterns, but its last three lines contain one of the most notorious instances of phonetic symbolism in English poetry — as we shall see in Tennyson’s poem. Likewise, statistical investigations can show only numerical correlations between sound and meaning; we shall try to show, in the same poem, how they interact on the fine-grained distinctive-feature and meaning-component level.

In addition, sound iconicity is not the only, and possibly not even the primary, means to set the tone of a poem. A large part of the tone is set by qualities abstracted from the meaning. It is possible that different poets rely to different degrees on sound symbolism, so even when a large corpus study fails to find a significant effect of sound symbolism, it does not mean that the phenomenon does not exist in any poem within the sample. Moreover, as we have seen in Miall’s (2001) finding with reference to Milton’s use of wide and narrow vowels in the description of heaven and hell, poets may change the rules of the game, drastically shifting from one kind of features to another, on which they rely in generating emotional qualities. This logic that we promote is different from the one guiding the mainstream view in the social sciences. While most studies in the social sciences (and possibly some scientific studies of literature) want to know whether the entire population has a certain property or not, we want to know whether the property exists somewhere within the population, even if it is not the most marked characteristic of the population.

To illustrate the importance of local sound pattern analysis over a global frequency analysis, consider excerpt 3, the complete text of Tennyson’s “Come down, O Maid”:

Come down, O maid, from yonder mountain height:
What pleasure lives in height (the shepherd sang)
In height and cold, the splendour of the hills?
But cease to move so near the Heavens, and cease
To glide a sunbeam by the blasted Pine,
To sit a star upon the sparkling spire;
And come, for Love is of the valley, come,
For Love is of the valley, come thou down
And find him; by the happy threshold, he,
Or hand in hand with Plenty in the maize,
Or red with spirited purple of the vats,
Or foxlike in the vine; nor cares to walk
With Death and Morning on the silver horns,
Nor wilt thou snare him in the white ravine,
Nor find him dropt upon the firths of ice,
That huddling slant in furrow-cloven falls
To roll the torrent out of dusky doors:
But follow; let the torrent dance thee down
To find him in the valley; let the wild
Lean-headed Eagles yelp alone, and leave
The monstrous ledges there to slope, and spill
Their thousand wreaths of dangling water-smoke,
That like a broken purpose waste in air:
So waste not thou; but come; for all the vales
Await thee; azure pillars of the hearth
Arise to thee; the children call, and I
Thy shepherd pipe, and sweet is every sound,
Sweeter thy voice, but every sound is sweet;
Myriads of rivulets hurrying thro' the lawn,
The moan of doves in immemorial elms,
And murmuring of innumerable bees.
To analyse the overall sound contents of the poem, we generated a phonetic transcription of the text using an online application (https://tophones.com), and calculated the frequencies of nasals and plosives using the data analysis software *Child Phonology Analyzer* (Gafni, 2015). We compared the obtained frequencies for those reported for the general English language in the early 20th century (Dewey, 1923), as close as possible to the time period of text.

Our analysis revealed that the poem contained 12% nasals and 17.5% plosives out of 823 phones (relative frequency ratio Nasal:Plosive = 0.69), compared to 11% nasals and 19% plosives reported for general English (relative frequency ratio Nasal:Plosive = 0.59; total 372,729 phones). Crucially, a $\chi^2$ test did not reveal a significant difference between the poem and the general language ($\chi^2(1) = 1.51, p = .22$). In other words, the analysis of phone frequency does not predict a particular emotion quality for the poem. This finding demonstrates our earlier claim that the sound content of texts is constrained by the arbitrary signs of the lexicon of the language.

Now, consider only the last two lines of the poem. The 50 phones in these lines contain 24% nasal consonants and 8% plosive consonants (relative frequency ratio Nasal:Plosive = 3). This distribution is significantly different from the general distribution in the English language ($\chi^2(1) = 9.95, p < .01$). Thus, we get markedly different results when analysing the sound contents of an entire poem compared to a shorter fragment. This finding, too, supports our argument in favour of a local, structural analysis of sound patterns. In what follows, we will apply this principle to the last two lines of the poem.\(^\text{10}\)

\(^{10}\) In fact, there are quite a few prominent local sound patterns in this poem; most arresting in this respect is the line

To sit a star upon the *sparkling* spire;
Or, consider the word pairs *spurted* purple and *Morning...horn*.  

The word ‘moan’ seems to be a perfect example of Auracher et al.’s (2011) generalization: It contains two nasal consonants and a long ‘dark’ back vowel, and it denotes a sound expressing a negative emotion, closely related to grief. On closer look, however, this appears to be not so evident. Obviously, ‘The moan of doves’ contains a metaphor; but metaphor for what? The Random House College Dictionary (1983) defines ‘moan’ as follows: “prolonged, low, inarticulate sound uttered from or as from physical or mental suffering.” The question is whether ‘moan’ is applied here to some ‘prolonged, low, inarticulate sound’ the nature of which must be inferred from the context, or to expressing some ‘physical or mental suffering.’

What we know about the mental state of doves when cooing and about the general mood of the stanza suggests that it is the ‘prolonged low sounds’ ingredient of moaning that activates the expressive potential of the nasals and the vowel. Since nasals are continuous, periodic, lowly encoded and of relatively low pitch, they are exceptionally well suited to suggest ‘prolonged low sounds’; the long back vowel reinforces this effect. This is reinforced by the next line too. ‘Murmur,’ in Lockwood and Dingemanse’s (2015) terms, “shows sensory sound-symbolism in that reduplication in the word is associated with a continuous meaning” (p. 3). The dictionary defines ‘murmur’ as “a low continuous background noise.” According to Wiseman and van Peer (2002) and Auracher et al. (2011), nasals express grief; according to Fónagy (1961), nasals tend to express tender emotions in general. In this case, there is no real contradiction between them, because grief is a tender emotion, only the latter is more general, including other kinds of emotions, such as love and pity as well, that is, both negative and positive emotion.

Now what do we know about prolonged low sounds in Tennyson’s stanza? It’s explicitly said that “sweet is every sound,/ Sweeter thy voice, but every sound is sweet” — the prolonged low sounds in the last two lines are pleasurable natural sounds underpinning the
pleasures of love, rather than woeful sounds. “The moan of the dove” is here a sensuous metaphor that suggests prolonged low natural sounds that happen to underpin some pleasurable, tender emotions, in harmony with Fónagy’s (1961) findings. This, however, does not exclude the possibility that in some other poem, the same phrase would suggest some physical or mental suffering, a tender emotion with some woeful taint underpinned by lowly encoded, periodic nasals with a ‘dark’ back vowel. Such is the nature of creativity in symbolic language.

Tennyson also uses in these lines the words ‘immemorial,’ ‘elms’ and ‘innumerable,’ and in the preceding line ‘lawn.’ The meaning of these words has nothing to do with either ‘prolonged low sounds’ or ‘physical or mental suffering’; in these words, the liquids and nasals and low back vowels are arbitrary signifiers of their meanings. But the poet, as well as his readers, may detect in the liquids and nasals and back vowels of these words, after the event, the features that generated an ‘iconic’ or expressive relationship between the liquids and nasals and the meanings in ‘moan’ and ‘murmur,’ reinforcing the iconic or expressive effect in those words. Here the effect is generated not by inherent iconicity (as could be inferred from some present-day studies), but by virtue of the infinite creativity and flexibility of symbolic language. By the same token, they generate a general psychological atmosphere spread over the two lines. As we have seen, the ‘physical or mental suffering’ ingredient in the meaning of ‘moan’ is cancelled by conflict with the context, in favour of pleasurable natural sounds — providing, again, an example of the creativity and unpredictability of symbolic language. In some other context, we may retain the ‘physical or mental suffering’ component, by deleting the <+human> transfer feature.
Sound-Symbolism and the Lexicon

Many scholars (e.g., Whissell, 1999) have noted that certain speech sounds are typically associated with certain positive or negative emotion words in the lexicon. Many articles quote Whissel with approval. They usually attribute this to a principle epitomized by Jakobson and Waugh (2002) as “inmost, natural similarity association between sound and meaning” (p. 182). Our theory explores this on a more fine-grained level, of phonetic and semantic feature interaction. Consider Aryani et al.’s (2018) finding that “analyses revealed crucial phonetic features potentially causing the effect of sound on meaning: For instance, words with short vowels, voiceless consonants, and hissing sibilants (as in ‘piss’) feel more arousing and negative” (p. 1).

How can voiceless plosives and short vowels suggest arousal and unpleasantness in obscenity? They don’t necessarily suggest arousal or unpleasantness; rather, the sharply defined boundaries of /p/ and /ɪ/ suggest ‘uncomplicatedness, directness’: straightforward violation of taboos. Note, however, that the violation of taboos exists (as well as the arousal and unpleasantness) in the meaning of piss, not in the speech sounds. The voiceless plosive only qualifies these qualities: It adds a sense of straightforwardness that renders the violation blatant, while the aperiodic sibilant adds a sense of roughness (in the case of piss, it also adds a sense of sound imitation: a stream of irregular, high sound stimuli). Likewise, in Aryani et al.’s (2018) other example, ‘terror’ in Edgar Allen Poe’s “The Bells,” the plosives may suggest the straightforwardness of the menace, and the sibilants, again, roughness, or even high-amplitude, menacing irregular noises. If Whissel (1999) associates plosives with unpleasant emotion terms, it is, perhaps, because some of those emotions may be characterized as straightforward.

In the introductory section of this article, we suggested that sound-symbolic lexical entries had been generated in a similar process to aesthetic instances, only fossilized into a lexical
entry. So, the boundary between aesthetic and pre-aesthetic is rather blurred. We have suggested that in our pre-aesthetic experiments, too, when we ask our subjects which one of two speech sounds is harder, smoother, more fuzzy-ended, we force them to enter the poetic mode. In certain lexical tasks, it is the meaning that forces the speaker-understander to enter the poetic mode, as with the verb, but not the noun, *swallow*.

Thus, as we have said, the relationship between speech sounds and emotional valence in the lexicon are epiphenomenal. Rather than inborn, we treat them as cognitive fossils. The central hypothesis of Tsur’s (2017b) book-length study is that conventions originate in a process in which, through repeated social transmission, cultural programs assume a good fit to the natural capacities and constraints of the human brain and cognitive system (D’Andrade, 1981). This is how such lexical regularities may have originated — with some entries, perhaps, already in the alleged protolanguage.

**Conclusions**

In this article, we discussed phonetic symbolism in poetry. The present argument can be summed up in one sentence: ‘Expressive sound effects are generated in poetry by the interaction of three components: semantic features, phonetic features and rhythm.’ Researchers who don’t make this distinction explicitly or implicitly may reach anomalous results.

We proposed a theory based on gestalt theory and linguistic feature analysis that accounts for apparently non-arbitrary sound–emotion relations based on structural similarities. In our conception, there is no direct sound–meaning iconicity, unless onomatopoeia refers to some elementary percept. Speech sounds can only suggest elementary percepts from which, when repeated in poetry, a general psychological atmosphere can be abstracted. This psychological atmosphere, in turn, can be individuated into specific emotions, by semantic feature addition
from the context. We have traced semantic feature transfer as an inherent function of our linguistic competence. The results of our stimulus–response experiment support this theory and demonstrate that different sound classes (nasals, voiced and voiceless plosives) are systematically associated with different emotional qualities (Gafni & Tsur, 2019). Furthermore, the same group of sounds (e.g., plosives) can convey, in the appropriate context, as different emotions as ‘joyful,’ ‘aggressive’ and ‘cautious,’ due to some structural features they share (sibilants may, at the same time, express some noisy, or some quiet, hushing quality [Hrushovski, 1980]). Aryani et al. (2018) relate the negative arousing power of obscene words as well as of terror to the same features of a combination of voiceless plosives and sibilants. We claim that sound symbolism can only support classification of emotions that is not based on valence (positive/negative, pleasant/unpleasant), since the physical object of sound does not have inherent valence, but they can foreground certain features of the emotion.

The proposed theory reconciles what may seem like conflicting findings in previous studies, by showing that they can be accounted for by the same structural principles. In addition, though large-scale investigation may yield significant results as in Fónagy’s (1961) study, it may miss significant local instances of sound symbolism, as we have demonstrated in Tennyson’s poem. Something like this may be responsible for Kraxenberger and Menninghaus’s (2016) conclusion that recurring speech sounds don’t seem to affect a poem’s emotional quality.

References


https://doi.org/10.1080/02699938908408075


lexical features of poetic texts-a case study using 57 Poems of H. M. Enzensberger.


