Phonetic symbolism: double-edgedness and aspect switching

Reuven Tsur and Chen Gafni

Abstract

This article proposes a structuralist-cognitive approach to phonetic symbolism that conceives of it as of a flexible process of feature abstraction, combination, and comparison. It is opposed to an approach that treats phonetic symbolism as “fixed relationships” between sound and meaning. We conceive of speech sounds as of bundles of acoustic, articulatory and phonological features that may generate a wide range of, sometimes conflicting, perceptual qualities. Different relevant qualities of a given speech sound are selected by the meaning across lexical items and poetic contexts. Importantly, speech sounds can suggest only elementary percepts, not complex meanings or emotions. Associations of speech sounds with specific meanings are achieved by extraction of abstract features from the speech sounds. These features, in turn, can be combined and contrasted with abstract features extracted from other sensory and mental objects (e.g. images, emotions). The potential of speech sounds to generate multiple perceptual qualities is termed double-edgedness, and we assume that the cognitive mechanism allowing us to attend to various aspects of the same speech sound is Wittgenstein’s aspect switching.

We discuss examples of phonetic symbolism in poetry, lexical items across languages, and laboratory experiments, and show how they can be accounted for in a unified way by our cognitive theory. Moreover, our theory can accommodate conflicting findings reported in the literature, e.g. regarding the emotional qualities associated with plosive and nasal consonants. Crucially, the theory is scalable and can account for general trends in poetry and lexical semantics, as well as more local poetic effects. Thus, the proposed theory can generate predictions that are testable both in laboratory experiments and in formal literary analysis.
Introduction

This article is derived from our book in progress on “Sound–Emotion Interaction in Poetry”, based on a group of chapters on phonetic symbolism. Many current studies in phonetic symbolism begin with a list of recent works on phonetic iconicity, that allegedly challenge de Saussure’s arbitrariness conception of the linguistic sign, followed by a series of experiments that establish a statistical relationship between phonemes (or groups of phonemes) and meanings or emotional qualities. In these works iconicity is assumed to be somehow given; very few of them go systematically into the structural relationship between the phonemes and the meaning or emotional qualities. In this practice, we strongly disagree with them on three issues.

First, we argue that literary phonetic symbolism does not challenge the arbitrariness hypothesis, but rather confirms it: sound-symbolic effects are attributed to verbal constructs after the event. Second, we believe that research based on stimulus–response questionnaires is, in itself, unsatisfactory. A cognitive approach should include a hypothesis, based on cognitive and linguistic research, to account for the structures in our mind that mediate between stimulus and response. Third, many of those stimulus–response studies take it for granted that rigorous quantitative analysis is the only valid way to evaluate empirical data. Furthermore, many such studies are interested only in global, gross effects (e.g. whether the frequencies of certain phoneme groups alone can account for the perception of poems as ‘happy’ or ‘sad’). We have pointed out a welter of incongruent statistical results in phonetic symbolism (Tsur & Gafni, forthcoming). The confusion in this domain originates in the neglect of a systematic theoretical framework. This neglect may manifest, among other things, in failure to treat sound–meaning relationships at a sufficiently fine-grained level, insufficient awareness of the complexity of the subject matter, as well as confusion whether sound-symbolic effects are motivated by acoustic, articulatory or phonological aspects of speech sounds.

The aim of this article is to present a comprehensive theoretical and analytic approach to sound–emotion relationships in poetry. At the basis of our investigation are earlier works by Benjamin Hrushovski (“interaction view”; 1980) and Iván Fónagy (“Communication in Poetry”; 1961), followed by Tsur’s work in the nineteen-eighties. Both Hrushovski and Fónagy revealed, by different methods, important relationships between speech sounds and their perceived effects.

In a structuralist theoretical discussion of sound–meaning relationships in poetry, Hrushovski explored the complexity of the subject, pointing out sources of complexity, of
which we shall discuss here only one: what we call the “double-edgedness” of sibilants: that in some contexts they may express noise of varying intensity, in some — a hushing quality. Alternatively, they may be “neutral”, not related to “meaning” or attitudes at all. Consider the following excerpts:

1. When to the sessions of sweet silent thought
   I summon up remembrance of things past,
   I sigh the lack of many a thing I sought
   And with old woes new wail my dear time’s waste
   (Shakespeare, “Sonnet 30”)

2. And the silken, sad, uncertain rustling of each purple curtain
   Thrilled me — filled me with fantastic terrors never felt before;
   So that now, to still the beating of my heart, I stood repeating,
   "'Tis some visitor entreating entrance at my chamber door —
   Some late visitor entreating entrance at my chamber door; —
   This it is and nothing more.”
   (E.A. Poe, “The Raven”)

3. [ . . ] the deep sea swell
   [ . . ] A current under sea
   Picked his bones in whispers,
   (T. S. Eliot, “The Wasteland,” IV)

4. And swell renews the salt savour of the sandy earth (“Ash Wednesday”)

In Excerpt 1, the sibilants have a hushing quality; in Excerpt 2, they imitate the rustling of the curtains. In Excerpt 3 “there is a transition from the powerful noise of a sea swell to the sound of whispers”. In Excerpt 4, in the “patterning of sibilants hardly a trace of either sound or silence remains” (Hrushovski, 1980: 41). Hrushovski (following Kreuzer, 1955) takes it for granted that in Excerpt 1 the sibilants have a hushing effect. Intuitively this seems right. He also argues that this hushing quality focuses the sound pattern on the meaning of the words “sweet silent”. But he does not explain what it is in the sibilants that invests them with this hushing quality. We shall explore, on a more fine-grained level, how they can express such opposite qualities.

In his “Communication in Poetry”, Fónagy examined the relative frequency of phonemes in six especially tender poems and six especially aggressive poems by four poets in three languages (Hungarian, German and French). Most phonemes occur with the same frequency as in general; but the nasals and liquid [m, n, l] occur more frequently in the tender poems in all three languages, whereas the voiceless plosives [k, t] occur more frequently in the
aggressive poems. [r] is aggressive in Hungarian and to some extent in German, but not in French. His method is used today frequently in research, but Fónagy’s use differs from the more recent uses in several respects. Most importantly, he did not assume some rigid iconic relationships between specific speech sounds and emotions where one may predict the emotional quality of the poem from the recurring speech sounds, as some present-day researchers do (e.g. Auracher, Albers, Zhai, Gareeva, & Stavniychuk, 2011). Rather, he probed into the possibility that there may be some significant correlation between pairs of opposite emotional qualities and pairs of opposite phonemes (or groups of phonemes). This was a ground-breaking idea at his time. He did not regard his findings as the “inmost, natural similarity association between sound and meaning” (Jakobson & Waugh, 2002: 182), but as possible examples for a principle.

Tsur, in his earlier works (e.g. Tsur, 1992), went one step further beyond Hrushovski’s and Fónagy’s works. For instance, Hrushovski defines expressive sound patterns as follows: “a certain tone or expressive quality abstracted from the sounds or sound combinations is perceived to represent a certain mood or tone abstracted from the domain of meaning”. Tsur asked the question, how can “a certain tone or expressive quality [be] abstracted from sounds or sound combinations”. In his answers, he relied on distinctive features, and on speech research done mainly at the Haskins Laboratories. The present article provides the answers to such questions as part of a comprehensive theory.

Cognitive theory

Structural model of speech sounds

In this article, we shall point out the complexity of phonetic symbolism at several levels, and propose a cognitive mechanism that may handle it. The basis for our theoretical account is a structural model of speech sounds: we assume that speech sounds are bundles of features, on the acoustic, articulatory and phonological levels, and that the various features may have different expressive potentials. Thus, we claim that the various phenomena related to phonetic symbolism result from the way speech sounds are coded and organized in our mental database. Here, we focus on the features of consonants.

Consonants can be abrupt: 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, 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.

The structural model described above organises consonants in our mental database in terms of acoustic features. We believe that it is the acoustic level that best accounts for the emotional and imitative aspects of sound-symbolic effects (see also: Aryani, Conrad, Schmidtke, & Jacobs, 2018; Knoeferle, Li, Maggioni, & Spence, 2017; Marks, 1975; Ohtake & Haryu, 2013; Ultan, 1978). Accordingly, we will use this organisation to account for evidence in the study of phonetic symbolism. Note that the proposed model relies on categorical acoustics features (e.g. abrupt/continuous). However, some studies explain phonetic symbolism in terms of continuous acoustic features, such as pitch and intensity (Aryani et al., 2018; Knoeferle et al., 2017). Here we assume a categorical acoustic theory of phonetic symbolism, but keep in mind that a continuous theory of phonetic symbolism can be perfectly consistent with ours.

We should note that there are additional theories of phonetic symbolism (see review in Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015). Like the above-mentioned acoustic theories, articulatory theories of phonetic symbolism assume that there is an iconic relation between sound and meaning. However, while acoustic theories assume a direct mapping of acoustic features, articulatory theories assume that iconicity stems from articulatory sensations, such as the size of the oral cavity (A. Fischer, 1999). Finally, phonological-distributional theories attribute phonetic symbolism to systematicity. According to such theories, phonological regularities in the lexicon cue non-arbitrary mappings between sound and meaning, e.g. different distributions of consonants in nouns and verbs (Monaghan, Christiansen, & Chater, 2007). Importantly, symbolic relations resulting from statistical regularities in the lexicon are language-specific (and, in a sense, arbitrary), although some of them can be grounded in perception or articulation. According to our conception, phonetic

---

1 The term ‘continuous’ refers to airflow in general. This should not be confused with the phonological feature [Continuant], which refers only to airflow in the oral cavity. Nasal consonants are continuous but not continuant.
symbolism is triggered by acoustic features, articulatory features, and phonological regularities, in descending order of importance.

The expressive potential of speech sounds: double-edgedness and aspect-switching

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 some 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, that 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.

One of the main questions in the study of phonetic symbolism is how the same speech sounds give rise to fundamentally different effects. We attribute this flexible affective potential of speech sounds to what we call double-edgedness: speech sounds have multiple features, each having its own expressive potential. In any given context, the expressive potential of one feature can be emphasized at the expense of the potentials associated with the other features. The cognitive mechanism to handle changing sound effects and relationships, we claim, is Wittgenstein’s (1976) aspect-switching. According to Wittgenstein, aspect switching is the capability of understanding the request “pronounce the word ‘till’ and mean it as a verb or a

---

2 We have borrowed the term “double-edgedness” from Ernst Kris, who speaks of the “double-edgedness” of the comic: a comedy about a cuckold may be very funny; but a husband in the audience, who suspects his wife is unfaithful, may find it offensive rather than funny. In poetry, double-edgedness is generated by changing contexts.

3 The following anecdote can illustrate our ability to switch between alternative perceptions of the same speech sound (in this case, liquid). When Tsur was on a sabbatical in Yale, he was robbed at gunpoint on the parking lot of the foreign scholars’ lodgings. When his wife told their Japanese neighbour the story, he was shocked and exclaimed: “Vat, viz a livorvel?!” Later, Tsur told this story to Professor Yehoshua Blau, president of the Academy of Hebrew Language, and added: If he can say consistently the wrong phoneme, it means that he does distinguish between them. He answered “It is, rather, your ear that made the distinction. He probably pronounced each time the same speech sound, somewhere between [l] and [r]. When you expected a [r], it sounded too much like a [l]; when you expected a [l] it sounded too much like a [r]”.
preposition, or the word ‘March’ and mean it as a month name or an imperative verb” (1976: 214e). In the present context, aspect-switching is the capability of attending to different features of the same phones. Here we will demonstrate the notions of double-edgedness and aspect-switching in relation to two classes of speech sounds: voiced plosives and voiceless fricatives ([f] and the sibilants ([s] and [§]).

We have argued that voiced plosives are ambiguous: they consist of an abrupt plosion plus voicing, that is continuous and periodic; consequently, they are double-edged. In the perceived quality of voiced plosives, the voicing stream may add resonance to the plosive element or a more massive presence. In physical terms, we define ‘resonant’ as ‘tending to reinforce or prolong sounds, especially by synchronous vibration; full-bodied, vibrant’; in phonetic terms, as ‘where intense precategorical auditory information lingers in short-term memory’. In the ‘tender’ context, the voiced stops [d] and [g] seem to have a fuller, richer, more resonant, more rounded body. The ‘tender’ context seems to increase the share of the periodic ‘voiced’ feature in them. By contrast, in the ‘angry’ context, they are perceived as more compact and abrupt unitary speech sound.

Hrushovski points out that sibilants may both induce a hushing psychological atmosphere, or may imitate noises of varying intensity. How can sibilants express hushing or noise? We claim that the double-edgedness of sibilants is related to different aspects of the same noises. Sibilants are continuous but not periodic, a stream of irregular noises. We propose to point out three different perceptual potentials in this structure: it is perceived as fluid, as fricative (rough) or as unpredictable (and as such, threatening). The tender or hushing quality of [s, §] may have to do with their fluidity aspect. The Merriam-Webster Collegiate Dictionary defines ‘fluid’ as “having particles that easily move and change their relative position without a separation of the mass; capable of flowing; subject to change or movement; characterized by a smooth easy style”. This physical description of liquid substances (e.g. hot pitch) draws attention to two apparently contradicting aspects: their smoothly-moving overall mass and their particles that are easily moving and relatively changing their position. This is a convenient metaphor for the contradictory effects of the voiceless sibilants (and of [f]). Language users switch between these two aspects, according to the context. When a mother hushes her baby by sounding a prolonged [§], they both attend to the smoothly-moving mass of the fluid §§§§§.

The “noisy” quality of voiceless sibilants and [f] is likely related to their fricative (or aperiodic) aspect. “Fricative” denotes a type of consonant made by the friction of breath in a

---

4 In Appendix A, we will also discuss double-edgedness of the trilled (“rolled”) [r].
narrow opening, producing a turbulent air flow. In “And the silken, sad, uncertain rustling of each purple curtain”, listeners attend to the randomly changing particles of the fricatives, imitating noise. Thus, different meanings may foreground a smooth or a turbulent aspect of fricatives.

**Encodedness of speech sounds and modes of speech perception**

So far, we seemed to have taken for granted that speech sounds have expressive potential and that listeners are able to switch between different potentials. Yet, we have not explained how these expressive potentials are perceived and how they are related to the primary use of speech sounds – conveying lexical meaning. In short, we claim that the two functions of speech sounds are related to different modes of auditory perception. Liberman and his colleagues (e.g. Liberman & Mattingly, 1985) distinguish between the speech mode and the non-speech mode. In the auditory, or non-speech mode, we hear, as in the sonar, the sound shape as shown by the machine. In the phonetic, or speech mode, listeners hear speech sounds categorically, i.e. as unitary phonemes. In this mode, speech sounds are typically encoded: the precategorical acoustic information that transmits the speech sounds is restructured into an abstract phonetic category and excluded from consciousness. Tsur (1992) assumes that there is a third, poetic mode of speech perception. In the “poetic” mode, some of the precategorical sound information reaches awareness from behind the unitary linguistic category. We claim that attending to this precategorical auditory information is what enables listeners to attribute emotional effects to speech sounds independently of semantic effects. Moreover, we claim that whether precategorical auditory information can reach awareness depends on the degree of encodedness of speech sounds.

Consonants are encoded to varying degrees, that is to say, in some consonants one may consciously perceive only a hard and fast category, not the precategorical auditory information that transmitted it. Voiceless plosives are “thoroughly encoded”, that is, no precategorical acoustic information reaches awareness (hence compact, tight, and solid in perception). Voiced plosives are also highly encoded. For example, if you ask “which one is acoustically higher, [ba], [da], or [ga]”, most people will answer that they don’t know what you are talking about.

However, in some consonants, some of the precategorical auditory information does reach awareness. If you ask which one is higher [s] or [ʃ] most people will easily answer that [s] is higher. This is because some of the precategorical sound information does reach awareness from behind the unitary speech category. In other words, voiceless sibilants are less encoded
than plosives. In sibilant continuants, more of the rich precategorical information is available than in most of the consonants, and is more apt to draw attention than in nasals or glides, for instance. This may explain how very different perceptual qualities can be associated with voiceless sibilants.

The foregoing discussion on encodedness can benefit from an interesting analogy from personality psychology. Psychologists of perception-and-personality put forward a series of personality dichotomies, such as levelers and sharpeners, or rigid and flexible: one pole tends to be more rigid, the other more flexible. Persons at the rigid pole tend to be intolerant of meaningless sense perceptions, as the following paragraph by Richard Ohmann may indicate:

The leveler is more anxious to categorize sensations and less willing to give up a category once he has established it. Red is red, and there’s an end on’t. He levels (suppresses) differences and emphasizes similarities in the interest of perceptual stability. For him the unique, unclassifiable sensation is particularly offensive, while the sharpener at least tolerates such anomalies, and may actually seek out ambiguity and variability of classification. (Ohmann, 1970: 231)

Briefly, categories afford greater perceptual stability; readiness to abandon oneself to unique, unclassifiable sensations suggests greater emotional receptiveness. Tender attitudes are characterized by openness to rich sensory information; rigid attitudes tend to cling to abstract categories. Going back to the double-edgedness of sibilants: as opposed to the rigid linguistic category, the stream of rich precategorical auditory information available in [s] has a flexible, soothing effect, and babies seem to be sensitive to it. Thus, the tender or hushing quality of [s, š] may have to do with their fluidity aspect, treated as unique, unclassifiable sensations in the stream of rich precategorical auditory information. The “noisy” quality of these sibilants springs from the aperiodic nature of the very same sensory information, a stream of irregular noises; we are switching from one aspect to another of the auditory stream.

**Phonetic symbolism in the lexicon and in poetry**

In this section, we will demonstrate how our cognitive model can account for cases of phonetic symbolism found in the lexicons of various languages and in literary pieces.

*Phonetic symbolism in the lexicon*

---

5 In the Rorschach test, for instance, responses based on colours and shadings (as opposed to shape responses) indicate emotional responsiveness.
The claim Tsur has elaborated in his 1992 book (Tsur, 1992) is that in different contexts, different potentials of the various features of the same sounds may be realized. Thus, for instance, the sibilants [s] and [ʃ] may have at some level of description features with noisy potentials, as well as features with hushing potentials. Evidence for these opposite potentials can be found cross-linguistically in lexical items related to both silence and noise. In such lexical items, the meaning picks out the appropriate meaning potential from a number of meaning potentials of its speech sounds. For example, in English, ‘silence’, ‘hush’, ‘still’, but also ‘rustle’, ‘shout’ and ‘scream’ contain sibilants; the German words for silence are Stille and Schweigen; in Hebrew, two words for silence contain sibilants, šeket, has; in Japanese, Shizukesa means something like ‘calmness’, ‘tranquillity’ or ‘stillness’; in Chinese, sūjing means ‘silence, solemnly silent, peaceful’, while shāshā and shēng mean ‘rustle’. In Hungarian, the word for silence is csend, beginning with the affricate [tʃ], csitt! means ‘whisht!’, kuss! means ‘shut up!’, ‘zaj’, ‘nesz’ mean noise, the verbs suhog, susog, zörg mean ‘rustle’, and pisszen in negative sentences suggests ‘slightest noise’ (‘does not make even the slightest noise’).

**Literary phonetic symbolism**

As we said in the introduction, many scholars take phonetic symbolism as counter evidence for the arbitrariness of the relation between sound and meaning. Our approach to the question of arbitrariness can be exemplified by the first line of Excerpt 2, repeated here.

2. And the silken, sad, uncertain rustling of each purple curtain
   Thrilled me — filled me with fantastic terrors never felt before;
   So that now, to still the beating of my heart, I stood repeating,
   “‘Tis some visitor entreating entrance at my chamber door —
   Some late visitor entreating entrance at my chamber door; —
   This it is and nothing more.”
   (E.A. Poe, “The Raven”)

The word “rustling” involves onomatopoeia proper. In the string “silken, sad, uncertain”, the alliterating speech sounds are arbitrary; only after the event they become non-arbitrary icons of the noise. It is only the unforeseeable combinations of symbolic language that afford the generation of this sound effect. It does not challenge de Saussure’s “arbitrariness”

---

conception, but rather, supports it. Yet, even in literary phonetic symbolism, it is not the case that “anything goes”. In the next paragraph, we will apply the proposed cognitive theory to account for what may seem like arbitrary use of phonetic symbolism in poetry.

The double-edgedness of voiced plosives can explain one of Fónagy’s most intriguing findings regarding the relative frequency of /g/ and /d/ in Victor Hugo’s and Paul Verlaine’s poems. /g/ occurs 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% in his tender poems, and 1.35% in his angry ones. 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.

There are two ways to account for this discrepancy between Hugo’s and Verlaine’s handling of the emotional quality of [d] and [g]. One possibility is that it is sheer arbitrary, idiosyncratic attribution of emotional qualities to the phonemes that has nothing to do with systematic phonetic symbolism: that anything goes, provided that it is statistically significant. Another possibility is that Hugo and Verlaine make use of the afore-said universal ambiguous structure of voiced plosives combined with aspect switching. In Wittgenstein’s terminology, we may rephrase the request “pronounce the phoneme /d/” as “pronounce the phoneme /d/, and attend to the massive presence bestowed on it by voicing, or to the resonance of voicing”. In all instances the pronunciation is the same; there is a mental shift. If you attend to the massive-presence aspect of /g/ or /d/, they may have a strong aggressive potential; if you attend to their resonant aspect, they may contribute to a tender quality. Obviously, Hugo and Verlaine applied the same cognitive mechanism to these voiced stops, but with a reverse focus. This finding supports our claim that sound symbolism relies on sometimes conflicting latent qualities, and the poet has freedom which one to realize. Obviously, the voiced plosives [d] and [g] have two conflicting potentials each, and the two poets realized them in opposing ways. Thus, the poets’ preferences are idiosyncratic, but rely on an objectively describable ambiguity.

**Phonetic symbolism in the laboratory**

In this section, we describe results of laboratory experiments related to the issue of phonetic symbolism. We discuss studies demonstrating the cognitive mechanisms assumed to underlie
phonetic symbolism, as well as studies directly related to phonetic symbolism. The results of these experiments will be discussed in relation to the cognitive model proposed in this article.

Experimental evidence for the cognitive model underlying phonetic symbolism

There is rich experimental evidence for aspect-switching in speech perception, namely that speakers are capable of switching between hearing a speech sound as a unitary phoneme or a phoneme plus precategorical information. Repp (1981) found that subjects could be trained to perceive some of the precategorical sensory information behind a s–ş continuum. In this study, subjects listened to pairs of SV stimuli, where S was a synthesised fricative along the s–ş continuum (7 artificial categories with different spectral properties), and V was a vowel ([a] or [u]) isolated from natural speech. The task was to decide whether the fricatives in a pair were identical or different, disregarding the vocalic context. It was found that most participants performed poorly at this task and their responses relied mostly on whether the vocalic contexts were similar or different across the stimuli in the pair. Such performance pattern was referred to as *categorical perception*, and it was attributed to the lack of ability to segregate the fricative noise from the vocalic context. However, after some training that included listening to the isolated fricatives, most participants improved significantly in the discrimination task in a vocalic context; that is, they were able to focus on the acoustic properties of the fricative noise. This pattern of performance was referred to as *noncategorical perception*. Thus, there is evidence that listeners are able to switch between modes (or aspects) of speech perception.

There is also compelling experimental evidence that listeners have subliminal access to precategorical auditory information in identifying voiced (but not voiceless) plosives. In a series of meticulously controlled experiments, Louis C. W. Pols (1986) presented subjects with spoken Dutch sentences or with vowel-plosive-vowel segments isolated from the spoken sentences. The recorded stimuli were modified in several ways, including deletion of the plosive burst and deletion of the transition between the plosive and one or both of the surrounding vowels (deletion means replacing the modified section with a silent interval of the same duration). The task was to identify the plosive out of several candidates. It was found that in natural sentences (as opposed to isolated VCV segments) the burst is *less* important than the vocalic formant transitions for the identification of voiced plosives (but not of voiceless plosives).

---

8 Although experimental studies have yet to demonstrate aspect-switching in plosives, evidence from other phoneme classes suggests that plosives are likely to show aspect-switching, as well.
plosives) (p. 149–150). Thus, the perception of the formant structure is not merely possible, but necessary for the identification of plosives.

We further investigated the idea that people have access to pre-categorical features of speech sounds in a pre-aesthetic experiment (Gafni & Tsur, forthcoming). We asked participants to read out loud pairs of consonant-vowel sequences (e.g. ma-ba) and compare the consonants on various bipolar perceptual scales (e.g. whether m sounds smoother and b sounds jerkier, or vice versa). Our experiment was designed specifically to test the hypothesis that voiced plosives are double-edged by contrasting them with voiceless plosives, on the one hand, and with nasals, on the other hand. Our hypothesis was largely supported. On the smoothness scale, we obtained a three-way contrast that was statistically significant: nasals were perceived smoother than voiced plosives that, in turn, were perceived as smoother than voiceless plosives. On other scales, we obtained only partial contrasts (possibly due to lack of statistical power): nasals were perceived as having fuzzier boundaries than voiced plosives that, in turn, were perceived as having fuzzier boundaries than voiceless plosives. However, the latter contrast was only ‘near significant’. In addition, we received two one-sided contrasts: first, voiced plosives were perceived as harder than nasals, but voiceless plosives were not perceived as harder than voiced plosives. Second, voiced plosives were perceived as thicker than voiceless plosives, but nasals were not perceived as thicker than voiced plosives. Whether these imperfect results reflect the true state-of-affairs or not, they clearly demonstrate that voiced plosives are perceptually ambiguous: in certain contexts, they contrast with nasals and, in others, with voiceless plosives.

Our experiment also yielded a puzzling finding. Voiced plosives were perceived as having more resonance than voiceless plosives, though the result was only near significant. This was well expected. However, contrary to expectation, voiced plosives were also perceived as having more resonance than nasals (the result fell short of statistical significance after correction for multiple comparisons). A simple explanation for this unexpected result is that the task was not clear to the participants. As a matter of fact, two participants commented that they had trouble evaluating resonance, and, in general, it became clear to us that many people don’t know what resonance means. However, there is also a possibility that the results are genuine, namely that voicing couples with plosion and endows voiced plosives with a

---

9 Formants are concentrations of overtones that uniquely identify vowels. Formant transitions are the rapid change in frequency of a formant for a vowel immediately before or after a consonant, and give information about the vowel and the consonant simultaneously.

10 A priori, nasals are expected to be judged as more resonant than their oral counterparts since, in the nasals, the nasal cavity vibrates, in addition to the sound articulated by the oral stop.
resonating quality, which can be perceived, at least in controlled experiments, out of context. This hypothesis is supported by experimental evidence that voiced plosives are perceived as larger than voiceless plosives cross-linguistically (Shinohara & Kawahara, 2016).

To conclude, there is experimental evidence supporting the cognitive mechanisms we claim to underlie phonetic symbolism, including both double-edgedness and aspect-switching. To be sure, these experiments do not prove that these mechanisms are directly involved in phonetic symbolism. However, they do demonstrate that participants have access to information underlying phonemic categories. We argue that having access to such precategorical information is the basis of phonetic symbolism.

Experimental evidence for sound symbolism

One of the most discussed examples of phonetic symbolism is that of sound-shape symbolism. Back in the nineteen twenties, Köhler (1929) took two nonsense words, takete and baluma, and asked people to match them with two nonsense figures, one with angular edges and one with rounded edges (see Figure 1). An overwhelming majority of respondents matched takete with the angular shape, baluma with the rounded shape. 11 Since Köhler’s study, there were many replications of this effect, including a study by Ramachandran & Hubbard (2001), which used the nonsense words bouba and kiki. So, this effect came to be called the ‘bouba/kiki’ effect.

![Figure 1. Shapes used in ‘bouba/kiki’ experiments (from Ramachandran & Hubbard, 2001).](image)

Recent studies have used more systematic manipulations in attempt to account for the ‘bouba/kiki’ effect. For example, participants in a study by Knoeferle et al. (2017) heard non-word sequences of consonant + vowel (e.g. /ba/, /ni/) and were asked to rate each sequence on a scale. One end of the scale indicated greater association with an angular shape and the other end indicated greater association with a rounded shape. It was found that when the stimulus

---

11 In the second edition of his book, Köhler (1947) changed baluma, for by now obvious reasons, to maluma.
contained a glide (w, y) or a liquid (l, r), participants tended to associate the stimulus with a rounded shape. This tendency was reduced for stimuli containing a voiceless fricative (f, s) or a nasal (m, n), followed by stimuli containing a voiced fricative (v, z) or a voiced plosive (b, d), followed by stimuli containing a voiceless plosive (p, t). The results are summarised in Figure 2.12

<table>
<thead>
<tr>
<th>Voiceless plosive</th>
<th>Voiced plosive, Voiced fricative</th>
<th>Nasal, Voiceless fricative</th>
<th>Glides, Liquids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angular</td>
<td></td>
<td></td>
<td>Rounded</td>
</tr>
</tbody>
</table>

**Figure 2.** Sound-shape matching (based on Knoeferle et al., 2017).

We can use the structural theory of speech sounds to account for these results. The left end of this figure is marked by [+ABRUPT], that is, [–CONTINUOUS, –PERIODIC]. This acoustic characteristic of voiceless plosives is analogous to the contour of angular shapes, which involves abrupt changes of direction. By contrast, the right end of Figure 2 is marked by [+CONTINUOUS, +PERIODIC]. This acoustic characteristic of glides and liquids is analogous to the smooth, gradually changing contour of rounded shapes. Thus, we can account for the tendency to associate voiceless plosives with an angular shape and glides and liquids with a rounded shape.

Let us now turn from the edges of the scale in Figure 2 towards the middle, ignoring fricatives for a moment. Nasals, like glides and liquids are continuous, periodic and voiced by default. Accordingly, they are associated more with a rounded shape than with an angular shape. It is not entirely clear why nasals would be rated lower on the scale than glides and liquids. One possibility is that another feature is at play here. For example, in articulatory terms, liquids are [+CONTINUANT], while nasals are [-CONTINUANT], since air flows continuously through the oral cavity in liquids but not in nasals. Thus, nasals might be considered more abrupt than glides and liquids, presumably since the nasal cavity is narrower and more disruptive for airflow than the oral cavity.

---

12 To be precise, all the effects of phonemic contrasts reported by Knoeferle et al. were calculated with glides as a baseline group. Thus, we only know for certain that liquids were NOT significantly different from glides while all the other classes WERE significantly different from glides. The distribution depicted in Figure 1 reflects our own interpretation of the results.
Moving further left on the scale, we see that voiced plosives are associated more with an angular shape than with a rounded shape. Like with voiceless plosives this can be attributed to their abrupt nature. But voiced plosives are ambiguous: they consist of an abrupt plosion and voicing, which is continuous and periodic. Thus, they are relatively abrupt, but not quite as voiceless plosives, and this difference is reflected in their rating on the shape scale.

We have replicated the results reported by Knoeferle et al. in an experiment that was designed specifically to test the contrast between voiceless plosives, voiced plosives, and nasals (Gafni & Tsur, forthcoming). In addition to establishing the effects of abruptness and voicing, we also found an effect of place-of-articulation. Labels containing labial consonants (b, p) were judged as more appropriate names for the rounded shape, compared to labels containing alveolar (d, t) or velar consonants (g, k). The results are illustrated in Figure 3. The source of this effect could be either acoustic or articulatory (Knoeferle et al., 2017). An articulatory account builds on analogy between lip rounding and a rounded shape. An acoustic account might be based on spectral differences among different places of articulation. A support for this view can be found in previous studies, e.g. “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/… [velar stops] display a stronger concentration of explosion [than labials and dentals]” (Jakobson & Waugh, 2002).

![Figure 3. The place of articulation sound-shape hierarchy (Gafni & Tsur, forthcoming)](image)

It is interesting to note that the basic demonstration of the bouba/kiki effect (Hung, Styles, & Hsieh, 2017; Ozturk, Krehm, & Vouloumanos, 2013; Ramachandran & Hubbard, 2001) contrasts a voiceless velar plosive (k) with a voiced bilabial plosive (b). The on-going discussion over the effect of place-of-articulation can explain why /b/ was preferred over other voiced plosives in these demonstrations (but note that this choice inadvertently introduces a confounding factor to the experiment). In addition, the effect of place-of-articulation might help explaining the difference between liquids and nasals found by Knoeferle et al. (if it was, indeed, significant). Both experiments included bilabial (m) and alveolar (n) nasal consonants

---

13 We did not test glides and fricatives in our experiment. In addition, in our experiment the non-word labels were presented in a written form rather than in a spoken form. Also, the labels we used were disyllabic rather than monosyllabic, and each label contained two identical consonants but two different vowels (e.g. *mamu, pupa*).
that were analysed as one group. In our experiment, \( m \) was rated higher (i.e. closer to the rounded shape) than \( n \) (mean ratings: 4.07 and 3.93, respectively, on a scale of 1-5), and \( l \) was rated slightly higher than \( m \) (mean rating: 4.09). Although the differences were not significant, it is possible that lip rounding (or its acoustic correlates) should be included in the structural model of phonetic symbolism.

At this point, let us reintroduce fricatives to our discussion. As can be seen in Figure 2, voiceless fricatives tended to be associated with a rounded shape more than with an angular shape. One possible explanation for this finding is that shape judgments are somehow affected by relative encodedness. As mentioned earlier, glides, liquids, nasals, but also voiceless sibilants are unencoded relative to plosives. In other words, rich precategorical auditory information is more accessible in the voiceless fricatives, liquids, nasals and glides than in plosives. Perception of this information seems to blur the sharp outlines of the categories, rendering these phonemes more appropriate to designate rounded than angular shapes (that is also one reason why they are perceived as ‘softer’).

However, if shape judgments depend on encodedness, one might ask, why voiced fricatives do not group with their voiceless counterparts. In the voiced fricatives, there are two parallel continuous streams, one periodic (voicing), and one aperiodic (friction); one might cautiously speculate that the two kinds of parallel streams blur each other and their perceived shape is less pronounced. This would suggest that the perceived roundness of speech sounds might be correlated with the relatively clear perception of the precategorical auditory information lingering in short-term memory.

**Resonance, co-articulation, and precategorical perception**

As we have said, resonating phones promote the perception of precategorical auditory information. Yet, there is evidence that the effect of resonance is modulated by context. For example, it was found that vowels in consonantal context are perceived more linguistically than isolated vowels (Rakerd, 1984); that is, precategorical sensory information can be better perceived in isolated vowels than in consonantal context. In other words, co-articulation seems to prevent access to precategorical auditory information. Some symbolist poets exploited this property of isolated speech sounds, to amplify the resonance of their poems. The French symbolist poet, Arthur Rimbaud, wrote in his sonnet “Voyelles”:

5. A noir, E blanc, I rouge, U vert, O bleu: voyelles
The Hungarian symbolist poet, Dezső Kosztolányi, wrote in a poem on his wife’s name, “Ilona”:

6. Csupa l, It’s all l
   csupa i, all i,
   csupa o, all o,
   csupa a. all a.

Thus, the examples above demonstrate that access to precategorical information, both in vowels and voiceless sibilants (and, probably, in liquids and nasals too), depends on isolating the vowels from the consonants, generating a resonating effect. Such a strategy would not work, however, with voiced plosives, because they cannot be pronounced without any vocalic context; in fact, formant transitions — that give information about plosives — give, by the same token, information about the adjacent vowels (speech researchers call this “dual transmission”). Nevertheless, the results of our experiment (Gafni & Tsur, forthcoming), demonstrated that, in minimal contexts (consonant-vowel sequences), voiced plosives can be perceived at least as resonant as nasals (though these results should be regarded with caution). We then thought to test whether voiced plosives are perceived differently in a reverberating or non-reverberating context, using an implicit task that does not require participants to have an explicit concept of resonance (see descriptions of such tasks in the next section). Consider Tennyson’s line and John Crowe Ransom’s rewriting of it:

7. And murmuring of innumerable bees
8. And murdering of innumerable beeves

There is general consensus that in the sound structure of Excerpt 7 the repeated nasals and liquids are perceived differently from Excerpt 8. In 8 “the euphony is destroyed […] we lose the echoic effect” (Abrams & Harpham, 2009). This is not only because in 7 there is one nasal more than in 8, but also because in 7 the meaning foregrounds the resonant quality of voicing, so that the nasals and the liquids have a fuller, richer, more resonant body. There is an intuition that even [b] is perceived as more resonant in Excerpt 7 than in Excerpt 8. To put it bluntly, in Excerpt 8 it sounds hard and compact as plosives are supposed to sound; in Excerpt 7 it drifts slightly toward the more resonant quality of [m]. This intuition is supported by our experiment. However, testing it explicitly is more difficult because the effect of the phonological context
(reverberating/non-reverberating) is confounded with the meaning of the line. Thus, if [b] is perceived as less resonant in 8, it could be due to the change of phonological context or the change of meaning.

**Phonetic symbolism, innateness, and the brain**

The Bouba/kiki effect (see earlier) has been demonstrated cross-culturally, even with Himba participants of Northern Namibia who had little exposure to Western cultural and environmental influences, and who do not use a written language (Bremner et al., 2013). Thus, the tendency to associate certain shapes with certain speech sounds is a cross-cultural phenomenon. Moreover, the effect has been demonstrated not only in forced-choice matching tasks, but also with implicit interference tasks. In a study by Westbury (2005), participants performed a lexical decision task on words and nonwords presented within curvy and spiky frames. Within each lexical category (word and nonword), some stimuli contained only stop consonants (e.g. toad and kide), some stimuli contained only continuous (e.g. moon and lole), and some stimuli that contained both stops and continuous sounds (e.g. flag and nuck). It was found that responses were faster for congruent shape-string pairs (continuous sounds in curvy shapes, plosives in spiky shapes) than for incongruent pairs.\(^\text{14}\) Using a masking technique, Hung, Styles, & Hsieh (2017) showed that the mapping for the *bouba/kiki* effect occurs prior to conscious awareness of the visual stimuli. Under continuous flash suppression, congruent stimuli (e.g. “kiki” inside a spiky shape) broke through to conscious awareness faster than incongruent stimuli. This was true even for participants who were trained to pair unfamiliar letters with auditory word forms. These results show that the effect was driven by the phonology, not the visual features, of the letters.

In another study, Ozturk, Krehm, & Vouloumanos (2013) presented 4-month-old infants with pairs of shapes and auditory stimuli. They found that the infants looked longer at the screen during trials with incongruent pairs (i.e. ‘bubu’ with an angular shape or ‘kiki’ with a curvy shape) than during trials with congruent pairs (i.e. ‘bubu’ with a curvy shape or ‘kiki’ with an angular shape). This finding, together with cross-cultural evidence, suggests that at least some aspect of sound-shape symbolism is pre-linguistic, perhaps even innate. But which

\(^{14}\) Note that stimuli in this study mixed voiceless and voiced plosives, affectively treating them as one group. As we have mentioned in several places, there is evidence that voiced plosives, and specifically /b/, are ambiguous. They can be grouped with voiceless plosives in certain contexts, and with sonorants, in other contexts. The Westbury study seems to force voiced plosives to be grouped with voiceless plosives.
aspect exactly? We claim that what is innate is not the symbolic relation per se, but rather the propensity to extract, abstract and compare abstract features from sensory stimuli. In what follows, we discuss other manifestations of this principle.

Roman Jakobson (1968) has shown that children’s acquisition of the phonological system of their mother tongue, the universal structure of phonological systems, as well as aphasic breakdowns are governed by the same principles based on abstracting and contrasting distinctive features. Assuming that we have similar capabilities with semantic and visual features, this would suggest that, whatever the brain processes involved, there is a universally available cognitive mechanism of abstracting, comparing and contrasting features that may be responsible for generating an indefinite number of unforeseeable sound-symbolic combinations.

One of the most striking pieces of evidence for our innate ability to contrast and combine abstract features in different domains can be seen in Synaesthesia. Synaesthesia is intersense perception, in which a person experiences sensations in one modality when a second modality is stimulated (Ramachandran & Hubbard, 2001). For example, a person may experience a specific colour whenever seeing a specific numeral (Galton, 1880). This trait seems to run in families, suggesting that it has a genetic and, thus, innate basis (Ward & Simner, 2005). In addition, neuroimaging studies have confirmed that synaesthesia is genuine, namely that the sensory experiences reported by synaesthetes are accompanied by activation in the related sensory cortical areas. Moreover, there is evidence that synaesthesia is related to an increase in structural connectivity among cortical areas involved in the synaesthetic experience (Rouw & Scholte, 2007).

Mature consciousness is based on the dissociation of the senses, but certain cognitive tasks, e.g. the bouba/kiki experiment still rely on synaesthetic mechanisms. Thus, although many adults do not experience synaesthesia on a regular basis, the cognitive mechanisms underlying synaesthesia seem to be present universally. Moreover, many languages have metaphors derived from the field of two senses, as in soft colour and warm sound — these are fossilized synaesthetic metaphors. The existence of such metaphors depend on, and attest to, the ability of individuals to extract and combine abstract features from different sensory and conceptual domains.

The use of such synaesthetic metaphors can be further extended in literary works via literary synaesthesia. Literary synaesthesia combines terms derived from different senses in unpredictable phrases, as in Keats’s “And taste the music of this vision pale”, where terms derived from three different senses are combined to suggest some elusive, nameless experience.
Phonetic symbolism in poetry is a special case of literary synaesthesia, where the imitative sound patterns are generated by unpredictable combinations of words. In Tennyson’s “And murmuring of innumerable bees” sound imitation is generated by repeating the speech sounds of *murmur* in *innumerable* — a word that has nothing to do with sounds.

Thus, it seems that writers have access to cognitive mechanisms, similar or identical to those involved in synaesthesia. Note, however, that despite the common surface similarity between “real-life” synaesthesia and literary synaesthesia, there is an important difference between the two phenomena. “Real-life” synaesthesia is automatic and involuntary, and involves stable and fixed cross-modal associations. By contrast, literary synaesthesia is a conscious, productive process that can give rise to an indefinite number of associations.

**Objections to the Foregoing Conception**

Auracher et al. (2011) criticized our conception that sound-symbolic effects are generated by an interaction between phonetic and semantic features as follows:

One plausible solution to cope with the contradictions between single studies is to follow Tsur’s (2006) and Miall’s (2001) hypothesis. Both authors conceive “speech sounds as of bundles of acoustic and articulatory features each of which may have certain (sometimes conflicting) combinational potentials, which may be activated, after the event, by certain meaning components” (Tsur, 2006, p. 905). However, this would not explain the high significance of the results within most studies. Why, as an example, would randomly chosen poems from different language families apply the same sound structure, expressing the same emotions, if the sound–meaning relation is content-dependent? (p. 22)

Part of our answer to this objection runs as follows: first, this question has a false presupposition, namely that poems that apply the same sound structure express the same emotions. It is false even with reference to the state of the art in our Western tradition. Auracher et al. found that sadness (and negative emotions in general) are best expressed by nasals, joy (and positive emotions in general) by plosives, whereas Fónagy (1961) found that tender emotions (that can be positive or negative) are typically associated with nasals, aggression (which is a negative emotion) with plosives.

Second, the “content-dependent” argument seems to us particularly invalid. Auracher et al. seem to assume that sounds express emotions irrespective of meaning (indeed, they hope to predict the emotion of a poem from the sound patterns). Furthermore, they seem to take Tsur’s (2006) claim as suggesting that sound-meaning relations in phonetic symbolism are arbitrary. Accordingly, they claim that if one takes the position that “sound–meaning relation is content-
dependent”, one should not expect sound effects to be translated from one language to another or from one culture to another. In their view, the fact that different studies on phonetic symbolism obtain significant results supports a notion of pre-defined sound-meaning relations. However, this is a misinterpretation of Tsur’s claim.

We don’t expect that “all aspects of [the word’s] meaning be deduced from its sounds”. We assume, rather, that phonetic symbolism can generate, at best, some vague psychological atmosphere, which the referent can individuate into a specific emotion by feature transfer from the meaning. This is part of our linguistic creativity. The results of such creativity may, indeed, fossilize into convention, e.g. a lexical entry; but in the bouba/kiki experiment, for instance, no convention is involved, both the words and their referents are unfamiliar, but participants do it spontaneously, and agree cross-culturally.

For example, let us consider the Japanese word ‘kirakira’. This word contains two elements that have iconic meaning crosslinguistically, reduplication and the vowel [i]. Crosslinguistic lexical surveys and laboratory experiments have demonstrated that reduplication can be associated with concepts such as ‘repetition’, ‘distribution’, and ‘intensification’ (O. Fischer, 2011; Imai & Kita, 2014)15, and the vowel [i] can be associated with concepts such as ‘smallness’, ‘brightness’, and ‘sharpness’ (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016; Lowrey & Shrum, 2007; Newman, 1933). Indeed, ‘kirakira’, which means “glittering”, “shows sensory sound-symbolism in that reduplication in the word is associated with a continuous meaning and the vowel [i] is associated with brightness” (Lockwood & Dingemanse, 2015: 3). We propose that the above-mentioned meanings, associated with the reduplication pattern and with the vowel [i], are all present in the sound pattern of ‘kirakira’, and that the semantics picks out two of them (i.e. ‘repetition’ and ‘brightness’) and eliminates the others (e.g. ‘smallness’, ‘sharpness’). Lockwood and Dingemanse suggest that “the vowel [i] is associated with brightness, but it also has conventionalized aspects in that not all aspects of its meaning can be deduced from its sounds”. We propose to handle the issue slightly differently. “Glittering” does not merely tilt the meaning in favour of brightness and repeated events (e.g. shimmering); it individuates the referent within the semantic field of light.

The stem “kiru”, in contrast to “kirakira”, means “to cut”. Here the meaning of /i/ is tilted toward sharpness and individuated as a specific, non-repeated action. Briefly, in phonetic symbolism, the sound structure of a nonsense word may pick out the appropriate unfamiliar referent in a forced choice; in a familiar lexical word, the meaning may pick out the appropriate

15 But also with completely opposite concepts such as diminutive and attenuative meanings.
meaning potential from a number of meaning potentials of its speech sounds. In a cross-cultural perspective, too, “the sound–meaning relation is content-dependent”. In Japanese, the principle works exactly as in English. Only in Japanese and Vietnamese there are formalized linguistic categories based on reduplication, whereas reduplications like “murmur” in English, or “rišruš” (rustling) in Hebrew, or “froufrou” in French, “ityeg-fityeg” (dangle repeatedly) in Hungarian are sporadic.

“The sound–meaning relation is content-dependent”, then, because the meaning components available in the semantic dimension of the text activate a subset of potentials derived from the universally available phonetic and acoustic features of the sound dimension. Characteristically, Auracher et al. rely on circumstantial evidence in the Stimulus–Response mode, “different cultures”, rather than cognitive evidence: whether semantic and phonetic information processing is similar or different in various cultures. This argument of Auracher et al.’s seems perfectly logical when treated on a wholesale, statistical level. But breaks down when treated at a more fine-grained, structural level.

The following comment has been made on our argument: “I don’t see how this claim [that the relation between sound and meaning is arbitrary] is consistent with the authors’ finding that ‘The structure of speech sounds determines their expressive potentials.”’ However, we don’t agree with this objection. The following observation by the same person is a clear support for our argument.

Neologisms and relatively recent creations like English “splat,” which have no etymologies of the usual kind, are of very minor interest in the study of literature, but they pose severe problems for the claim that meaning has nothing to do with linguistic sound. According to Fidler (2014: 229), this is expected, since OpEs (onomatopoetic expressions) provide “the simplest way to express effectively maximum information with minimal effort and the simplest way for the hearer to process the information intuitively and with minimum effort.”

How do we produce and understand neologisms? How does “the simplest way to express effectively maximum information with minimal effort and the simplest way for the hearer to process the information intuitively and with minimum effort” work? We argue that speech sounds are part of a phonological system, and are bundles of phonological, acoustic and articulatory features. According to Jakobson (1968), the child’s language has two stages: first a babbling period, in which the child experiments with the articulation and the acoustic qualities of the speech sounds, which are emotionally charged; then the arbitrary referential use of the same speech sounds, governed by the will. Mastering the arbitrary referential use is
perceptually governed, by maximum contrast. The first syllable children in all languages master for referential use is Pa, which consists in maximum opening and maximum closure of the lips; from this they derive “papa”. The reduplication indicates here that this time the articulation of speech sounds is no mere pleasurable experimentation with speech sounds, but used for arbitrary reference.

Then they master the contrast between labial and dental consonants, obtaining “tata”. Then they master the contrast between oral and nasal consonants, obtaining “mama” and “nana”, and so forth. The process is not iconic, but perceptual and articulatory; meaning is attached in an arbitrary manner. There is no iconic relationship between “papa” and the male parent, or “mama” and the female parent, the terms and references are derived from the system. But then, perhaps, children have an intuition that mothers are contrasted to fathers in being softer, assigning nasals to mothers and plosives to fathers rather than vice versa.

We argue that in adult language, speech sounds, as part of the phonological system, serve for arbitrary reference. The acoustic and articulatory features have a wide range of incongruent, dormant expressive potentials. When a label consisting of a sequence of speech sounds is arbitrarily attached to a meaning, it may (or may not) activate, after the event, some of the dormant expressive potentials. In the word “splat” and “splash” meaning activates certain expressive potentials of the speech sounds; in “split” or “plush” it does not. All language users have access to those dormant acoustic and articulatory features, so that they may activate them creatively and differentially. Sibilants have both hushing and noisy features; that is why they may abound, universally, both in words meaning noises of varying amplitude, or silence. That is why “Onomatopoetic expressions provide ‘the simplest way to express effectively maximum information with minimal effort and the simplest way for the hearer to process the information intuitively and with minimum effort.’” The alleged protolanguage is said to have been entirely onomatopoetic; symbolic language is typically arbitrary. Computer simulations show that iconicity is favoured in small systems, but arbitrariness has greater advantages as the system expands (Gasser, 2004). There is experimental evidence that iconicity facilitates for children the acquisition of a small, elementary vocabulary, but it obstructs the acquisition of a large and versatile vocabulary (Imai & Kita, 2014). Onomatopoetic words have two characteristics. First, their referents are sounds, so that the speech sounds may directly imitate them; second, in onomatopoetic words the fluid creative process described above has fossilized into a dictionary entry.
Future Research

Some of our experimental results were indecisive, owing to insufficient number of participants, or participants’ difficulty to understand such tasks as resonance judgment. We intend to repeat some of our experiments with a greater number of participants, and using interference tasks, where participants need not understand such elusive notions.

We assume that sound-symbolic and metaphoric productivity are driven by a homogeneous set of principles. Both are based on unpredictable feature interaction — one of phonetic and semantic features, the other of conflicting semantic features. We intend to explore this similarity.

Another future research plan aims to improve experimental methodology for literary studies, by exploiting new technologies. In recent years there has been an increase of literary studies employing experimental methods from social sciences, such as stimulus—response questionnaires. Such studies typically ask participants to evaluate entities on various pre-defined scales (e.g., rate how sad a poem is). Pre-defined evaluation scales are useful for answering specific research questions. However, such scales might be less suitable for evaluating the effects of aesthetic objects, since they can capture only a limited portion of the subjective experience and thus may predetermine the evaluated scales, missing important aspects of the aesthetic effect. Fortunately, recent advances in methods of natural language processing allow for sophisticated automatic analysis of texts. Such methods can be incorporated in experimental studies using, open-ended questions to evaluate aesthetic objects. Thus, instead of asking participants to rate poems according to pre-defined evaluative terms, the researcher can ask participants to provide their impression of the poem. Subsequently, using automatic text analysis methods, one can obtain a detailed, unbiased profile of the subjective experience. We plan to investigate the utility of such methods for analyzing responses to poetic text.

To Conclude

In this article, we reviewed three types of evidence for phonetic symbolism. First, certain types of phones tend to correlate with certain meanings cross-linguistically (e.g. the frequent occurrence of sibilants in word related to silence and to noise). Second, laboratory experiments demonstrate that participants tend to form consistent associations between speech sounds and concepts from various domains (e.g. made-up names assigned to rounded shapes tend to contain nasal and liquid consonants, while names assigned to angular shapes tend to contain
voiceless plosives). Third, various literary devices (e.g. sound repetition) use sound patterns, detached from semantics, to induce a general psychological atmosphere in poems.

We claim that the various cases of phonetic symbolism can be accounted for in a unified manner via a structuralist-cognitive theory. Central to this theory is the innate, general-purpose human capacity to extract abstract features from sensory objects (e.g. sounds, shapes). This capacity, together with our ability to combine and compare abstract features, allows us to attribute meaning potentials to speech sounds based on structural similarities to various objects. Importantly, the cognitive model does not determine sound-meaning associations, but rather constrain them. A given phone can have multiple, sometimes conflicting, meaning potentials, such that different aspects of the same phone are highlighted across lexical items and semantic contexts.

We referred to the multiple meaning potentials of speech sounds as *double-edgedness*, and proposed that people can shift their attention from one potential meaning to another via the cognitive mechanism of *aspect-switching*. We described some empirical results supporting the ideas of double-edgedness and aspect-switching. Moreover, we claim that these mechanisms can account for conflicting, or seemingly arbitrary, findings in the literature, such as the conflicting associations of nasals and plosives reported by Fónagy (1961) and Auracher et al. (2011).

The double-edgedness of voiced plosives was of special interest to us. Previous studies, as well as our own experiments, demonstrate that voiced plosives can behave like their voiceless counterparts in certain contexts, while in other contexts they behave more like sonorous consonants. Our structural analysis and experiments suggested that voiced plosives are double-edged. Some of these experimental results, however, may be due to participants’ misunderstanding of the task. More reliable is the fact that voiced plosives yielded significant results in experiments in which they were opposed to continuous phonemes, but also in experiments in which they were opposed to voiceless plosives. Likewise, a structural analysis showed, how [d] and [g] in poems by Verlaine and Hugo could be used to opposite effects. All this suggests that double-edgedness and aspect-switching may have psychological reality. However, more research is needed to investigate the cognitive mechanisms of double-edgedness and aspect-switching.

---

16 Note that Westbury (2005) puts voiced plosives in one bin with voiceless plosives, contrasting them to continuants, whereas Hung et al. (2017) and Ozturk et al. (2013) contrast the voiced plosive [b] with the voiceless plosive [k]. Köhler too puts [b] in one bin with [l] and [m] in “baluma”, contrasting them with the voiceless plosives in “takete”
Appendix A: Double-edgedness of the “rolled” [r]

Across languages, rolled [r] is frequently associated with noise. Many linguists believe that in some time in the past r was rolled in English, French, and Hebrew. English, French and some German dialects have lost their rolled [r] but, as the following examples witness, their lexicons still retain them as onomatopoeia (usually in combination with voiceless fricatives) for noise: In English, “rustling” and “whisper” contain /s/ and /t/. In French, bruit, froufrou, crissement, and susurrer mean “noise”, “rustle”, “screech”, “to rustle, whisper”, respectively. In German, knistern and rascheln mean “rustle, rustling”. In Hebrew, raʕaš and rišruš (in phonetic transcription) mean “noise” and “rustle”, respectively. sarasara in Japanese means “murmuring, rustling” (in both cases, reduplication indicates sound symbolism). Iván Fónagy writes to this effect:

Selon une tradition qui remonte à l’antiquité, le r est associé […] au combat. […] La distribution de fréquence des phonèmes dans les poèmes appartenant à deux populations sémantiques différentes, dans les poèmes belliqueux d’une part, dans les poèmes idylliques d’autre part, reflète en effet une forte tendance à associer les r à la violence […]. Le r français déjà «affaibli» à l’époque de Victor Hugo, n’est plus associé à la violence […]. Cette dureté de r apical semble qualitativement différente de la dureté des occlusives sourdes […]. Les R «sont les vrayes lettres Héroïques», écrit Ronsard dans la préface de la Franciade (1587, Œuvres complètes, Paris, 1914, VII, p. 93). (Fónagy, 1983: 96)

According to a tradition that goes back to antiquity, the r is associated […] with combat. […] The distribution of the frequency of phonemes in the poems belong to two semantic populations the bellicose poems on the one hand, and the idyllic poems on the other, reflects, as a matter of fact, a strong tendency to associate the rs with violence […] French r, already “weakened” by Victor Hugo’s time, is no longer associated with violence […]. This hardness of the apical r seems to be qualitatively different from the hardness of voiceless stops […] The Rs “are the true Heroic letters” wrote Ronsard in the preface to la Franciade (1587, Œuvres complètes, Paris, 1914, VII, p. 93). (Fónagy, 1983: 96)

Thus, [r] serves as onomatopoeia for noise as in roar, rustle. Yet, in the same languages, [r] may have a soothing effect as in Lili Marleen. Our cognitive theory of phonetic symbolism can account for this affective duality of [r]. The rolled [r] is periodic and multiply interrupted, thus, it is double-edged. Its abrupt nature makes it suitable for denoting noisy and hard qualities, but, in other contexts, its periodicity can induce a tender effect.

---


[18] It is unknown when exactly the quality of [r] had changed in these languages. Therefore, the ongoing discussion is somewhat speculative. For that reason, we decided to bring the analysis in an appendix.
Appendix B: Articulatory aspect switching

In this article, we discussed the mechanism of aspect switching in the perception of voiced plosives. There is another theoretical possibility, namely that the pronunciation itself changes with context, emphasising either the plosion or the voicing. Alternatively, one may speculate that perhaps different acoustic cues for phonological voicing are at work here. There are several acoustic cues for a voiced stop, usually used in conjunction:

“In distinguishing between voiced and voiceless plosives, the exact moment at which periodicity begins is among the cues used by the listener. […] and] the distinction between post-vocalic voiced and voiceless sounds is carried very largely by the relative duration of the vocalic and the consonantal parts of the syllable; in /biːt/ (beat) the vocalic part is relatively short and the interruption caused by the consonant is long, while in /biːd/ (bead) the reverse is the case” (Fry, 1970: 36).

Assuming that both kinds of cues work in conjunction, speakers might, involuntarily, emphasize one or the other cue, voice onset time or vowel length. As a rule, there is trade-off between competing acoustic cues for phonemes.19

With that said, note that although it is possible that the switch between alternative qualities of voiced stops is done by modifying their pronunciation, there is no evidence that the poets intended their poems to be read with such an overt articulatory aspect-switching, and there is no empirical evidence that readers actually do read the poems this way.

References


19 “While a listener typically perceives only a single change—viz., one of phonetic category—the physical changes that led to this unitary percept can only be described in form of a list with multiple entries. […] If one cue in such an ensemble is changed to favor category B, another cue can be modified to favor category A, so that the phonetic percept remains unchanged.” (Repp, 1983: 342)


Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language
acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369.* https://doi.org/10.1098/rstb.2013.0298


