Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pragmatics 38 (2006) 905–924
journal of PRAGMATICS
www.elsevier.com/locate/pragma

Size–sound symbolism revisited

Reuven Tsur

Tel Aviv University, Professor Emeritus, Hebrew Literature, The Cognitive Poetics Project, 69978 Tel Aviv, Israel

Received 15 November 2003; received in revised form 3 February 2004; accepted 1 December 2005

Abstract

Why do we perceive bass voices as ''thick''? Owing to a ''mediated association'' with ''thick people and animals [who] are usually loud and resonant'', or owing to some ''subtle inter-sensory quality'' found in thick things and bass voices? The present paper rejects the former possibility and endorses the latter. As to speech-sound symbolism, I account for it with reference to two aspects: phonetic features and precategorial information. I conceive of 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. Speech is transmitted by sound waves; but while speech categories are consciously perceived, the rich precategorial auditory information that transmitted them is excluded from awareness. I assume that intuitions regarding perceptual and emotional qualities of speech sounds are prompted by rich precategorial auditory information that subliminally reaches awareness in spite of all. Examining a sample of 136 languages, Russell Ultan (1978) pointed out that in a wide range of cultures high front vowels are typically perceived as small or denoting small things, whereas low back vowels are typically perceived as big or denoting large things. ''Since high front vowels reflect proportionately higher second formant frequencies, [. . .] there appears a correspondence between a feature of high frequency (=short wavelength in physical terms) and the category of small size''. Gérard Diffloth (1994) provides a counterexample: There is a word class in a Vietnamese dialect in which ''high is big'', and ''low is small''. Though I sympathise with quite a few of Diffloth's generalisations, I must conclude that he arrives at his contradictory findings by changing the rules of the game. The two researchers mean different things by the same words. Speaking of ''high'' and ''low'', Ultan means relative formant frequency; Diffloth means articulatory location. The ''height'' of the place of articulation of the vowel is in an inverse relation to the frequency of its first formant. ''High'' articulatory location is synonymous with ''low'' first formant frequency. So, in both instances ''high frequency is small'' and ''low frequency is big''. They differ, then, in that while Ultan's intercultural sample directs attention to the frequency of the second formant, Diffloth's Vietnamese word class focusses on the frequency of the first formant. As far as the size–vowel symbolism is concerned, the convincing counterexamples are still to be adduced.

Keywords: Speech perception; Sound symbolism; Synaesthesia; Aspect-switching; Frequency code; Perceived qualities; Mediated association; Phonetic features; Meaning components; Categorial perception; Precategorial auditory information

E-mail address: tsurxx@post.tau.ac.il.
URL: http://www.tau.ac.il/~tsurxx/.

0378-2166/$ – see front matter
doi:10.1016/j.pragma.2005.12.002

1. Preliminary

In many of my writings I have argued that poetic images have no fixed predetermined meanings. In my 1992 book *What Makes Sound Patterns Expressive?—The Poetic Mode of Speech Perception* (originally published in 1987) I propounded the view that speech sounds do not have fixed predetermined symbolic values either.[1] Poetic images as well as speech sounds are clusters of features, each of which may serve as ground for some combinational potential. The resulting combinations of images and speech sounds give rise to figurative meanings and sound symbolism. Unforeseen contexts may actualise unforeseen potentials of images and speech sounds. Language users may shift attention from one potential to another in the same speech sound or poetic image, and realise new figurative meanings and sound-symbolic qualities. Thus, the handling of figurative language and sound symbolism in poetry is governed by a set of homogeneous principles. The acquisition and use of language require considerable creativity. This creativity is heightened and turned to an aesthetic end in the writing and understanding of poetry. In my writings I have explored the sources of these potentials, and how human intuition handles them in generating poetic qualities.

In this way, a sophisticated interplay between sound and meaning is generated. Relevant features can be multiplied indefinitely, and one may discover unexpected phonetic or phonological features. In my 2001 paper ''Onomatopoeia: Cuckoo-Language and Tick-Tocking: The Constraints of Semiotic Systems'' I consider a minimal pair that can illustrate this. In Hebrew, *mətaktek* means ''ticktocking''; we attend to the repeated voiceless plosives and perceive the word as onomatopoeic. *mətaktak*, by contrast, means ''sweetish'', derived from *matok* (sweet). In Hebrew, the repetition of the last syllable is lexicalized, suggesting ''somewhat (sweet)''. A wide range of such ''moderate'' adjectives can be derived in this way from ''main-entry'' adjectives: *ḥamaşmaş* (sourish) from *hamuş* (sour), *adamdam* (reddish) from *adom* (red), *yərakrak* (greenish) from *yarok* (green), and so forth. Hebrew slang even derives *gəvarbar* (''somewhat man'') from *gɛvɛr* (man). The meaning directs our attention to this redoubling of the syllable, and we attend away from the acoustic features of the specific consonants. Benjamin Hrushovski (1968) pointed out that the sibilants have different (even opposite) effects in ''When to the sessions of sweet silent thought / I summon up remembrance of things past'' and in ''And the silken, sad uncertain rustling of each purple curtain''. In my book I explore the different aspects of the sibilants that may generate such conflicting effects. In the former quote, meaning components related to ''silent'' activate one set of aspects; in the latter, meaning components related to ''rustle'' activate another set. This is what Wittgenstein (1967:194) called ''aspect-switching''.

I wish to point out an additional issue, crucial for an understanding of how sound symbolism works. The Haskins Laboratories researchers distinguish between a speech mode and a nonspeech mode of aural perception, which follow different paths in the neural system. In the nonspeech mode we listen to a stream of auditory information in which the shape of what is perceived is similar to the shape of the acoustic signal; in the speech mode we ''attend away'' *from* the acoustic signal *to* the combination of muscular acts that seem to have produced it; and *from* these elementary movements *away to* their joint purpose, the abstract phoneme sequence. In this mode, all the rich precategorial sensory information is shut out from awareness. In verbal communication it is the abstract phoneme that counts, not the precategorial sound stream or the articulatory gestures that led to the abstract category. There is, however, experimental literature that gives evidence that some of the rich precategorial sensory information *is* subliminally

[1] The sound files for this paper are available online on the web: http://www.tau.ac.il/∼tsurxx/CreakyFolder/Creaky.html.

perceived. I have claimed that there is also a ''poetic mode'' of speech perception, in which emotional and perceptual qualities are generated when the precategorial auditory information is available for combination with meaning components. In the Hebrew word for ''ticktocking'' the meaning directs attention to the sensory information underlying the voiceless plosives; in the Hebrew word for ''sweetish'' it directs attention away, to an abstract lexical model.

The present paper was prompted by three chapters, by John Ohala, Eugene Morton, and Gérard Diffloth, in the mind-expanding collection of essays *Sound Symbolism* (Hinton, Nichols, Ohala (Eds.), 1994). In the light of these essays I will recapitulate two issues from the chapter ''Some Spatial and Tactile Metaphors for Sounds'' of my above-mentioned book: the relation between sound frequency and the size of the body that produced it; and the relation between ''high'' and ''low'' vowels and the suggested size of their referents. The former may account for the rise of certain crucial potentials active in the latter.

2. Phlogiston and precategorial information

An anonymous reviewer criticised my use of ''precategorial auditory information'' as follows. ''The construction of an entity called ''precategorial auditory information'' sounds to me like phlogiston reborn in linguistics: auditory information that is perceived without semantic understanding, yet somehow convey semantic understanding—assuming I've grasped the principle''.

The phlogiston comparison is delightful, but totally unfounded. So, to avoid such misunderstanding, I am going to spell out at some length what I mean by ''precategorial auditory information''. I have taken the term from Al Liberman and his colleagues. Speech sounds can be uniquely identified by acoustic energy concentrations at varying frequencies, called formants. The lowest is called F_1, the second lowest F_2, and so on. (Intonation pitch is F_0.) These can be turned by a ''spectrograph'' into patches of light and shade (or colours) called ''spectrograms''. Fig. 1 presents the spectrogram of the syllables /ba/, /da/ and /ga/, uttered by me.

From the hand painted spectrogram in Fig. 2 the syllables /di/ and /du/ can be synthesised on a machine called ''pattern playback''. The parallel horizontal first and second formants of Fig. 2 represent the vowels /i/ and /u/; the encircled portions preceding them represent rapid changes of frequency called formant transitions, in this case the second formant transitions. They give information about the vowel and the preceding consonant /d/ at the same time. This is called ''parallel transmission''. The syllables ''ba'', ''da'' and ''ga'' differ only in the second formant transition's onset frequency. There is, then, no resemblance between the shape of the perceived speech category and that of the precategorial acoustic signal that carries it. This is called ''categorial perception''. An illuminating instance of categorial perception has been demonstrated with reference to Fig. 3, where the pitch continuum between /ba/, /da/ and /ga/ has been divided into fourteen equal steps instead of three. Listeners discriminate the same difference more accurately near the category boundaries than within the categories. In the speech mode we hear only the unitary speech category, but not the sensory information represented by the spectrograms.

In certain artificial laboratory conditions one can hear directly the precategorial auditory information. So, as a second step, the reader may listen online to the sequence of syllables represented in Fig. 3, and to the sequence of isolated second formants from an unpublished demo tape by Terry Halwes. See whether you can hear a gradual change between the steps, or a sudden change from /ba/ to /da/ to /ga/. Halwes then isolates the second formant transition, that piece of

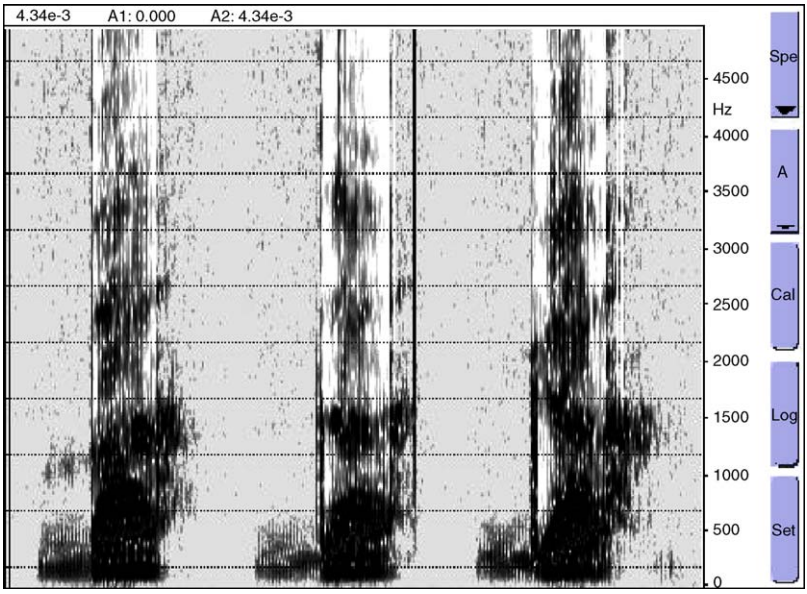

Fig. 1. Spectrogram of the syllables /ba/, /da/ and /ga/, uttered by myself.

sound which differs across the series, so as to make it possible to listen to just those sounds alone. Most people who listen to that series of chirps report hearing what we would expect, judging from the appearance of the formant transition: upward glides, and falling whistles displaying a gradual change from one to the next. The perception of the former series illustrates the speech mode, of the latter series—the non-speech mode.

As a third step, I am going to quote a substantial section from my 1992 book that presents some evidence that in some ''mysterious'' way the ''precategorial auditory information'' *can* be

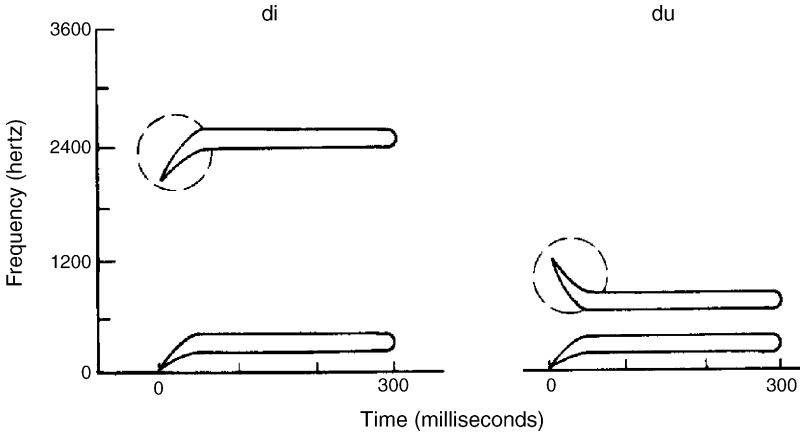

Fig. 2. Simplified hand-painted spectrogram from which the syllables /di/ and /du/ can be synthesised on a pattern playback.

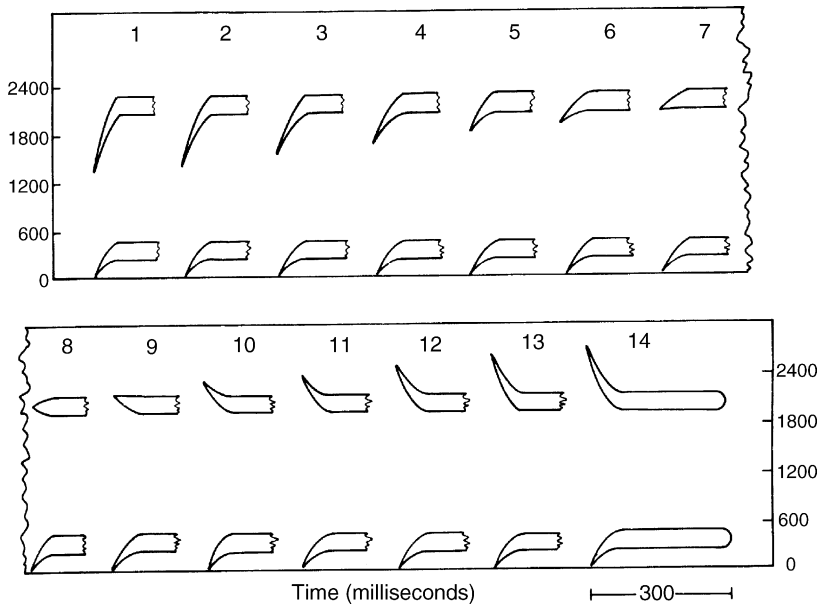

Fig. 3. Hand-painted spectrograms of the syllables ba, da, ga. The ba–da–ga pitch continuum of F_2 is divided into 14 steps instead of three. The two parallel regions of black indicate regions of energy concentration, F_1 and F_2. Notice that the onset frequency of F_2 of da is higher than that of ba; and the onset frequency of F_2 of ga is higher than that of da. Only stimulus 14 represents its full duration.

heard while being tuned to the speech categories. One such piece of evidence is provided by Liberman et al., who describe an experiment by T. Rand:

> To one ear he presented all the first formant, including the transitions, together with the steady-state parts of the second and third formants; when presented alone, these patterns sound vaguely like /da/. To the other ear, with proper time relationships carefully preserved, were presented the 50-msec second-formant and third-formant transitions; alone, these sound like the chirps we have referred to before. But when these patterns were presented together—that is, dichotically—listeners clearly heard /ba/, /da/ or /ga/ (depending on the nature of the second-formant and third-formant transitions) in one ear and, simultaneously, nonspeech chirps in the other. Thus, it appears that the same acoustic events—the second-formant or third-formant transitions—can be processed simultaneously as speech and nonspeech. We should suppose, then, that the incoming signal goes indiscriminately to speech and nonspeech processors. If the speech processors succeed in extracting phonetic features, then the signal is speech; if they fail, the signal is processed only as nonspeech.

I have arrived at the notion of "the poetic mode" on the basis of introspection and certain thought experiments, back in 1980. However, I have received massive corroboration to my speculations that nonphonetic acoustic information may be available in the speech mode too, when Repp's comprehensive survey of research on categorial perception was published in 1984. Thus, for instance, speaking of the within-categorial steps represented in Fig. 3,

> Liberman et al., (1957) were able to generate a fair prediction of discrimination performance from known labeling probabilities; however, performance was somewhat

> better than predicted, suggesting that the subjects did have some additional stimulus information available (Repp, 1984:245).

What is more, people appear to be capable of switching modes, by using different listening strategies. Fricative stimuli seem to be especially suited for the application of different strategies, so that they may be perceived fairly categorically in one situation but continuously in another (ibid., 287). Repp has investigated the possibility that with fricatives, for instance, little training would be necessary for acoustic discrimination of within-category differences. He employed a /s/-/ʃ/ continuum, followed by a vocalic context. The success of his procedure

> together with the introspections of the experienced listeners, suggested that the skill involved lay in perceptually segregating the noise from its vocalic context, which then made it possible to attend to its ''pitch''. Without this segregation, the phonetic percept was dominant. Once the auditory strategy has been acquired, it is possible to switch back and forth between auditory and phonetic modes of listening, and it seems likely [...] that both strategies could be pursued simultaneously (or in very rapid succession) without any loss of accuracy. These results provide good evidence for the existence of two alternative modes of perception, phonetic and auditory—a distinction supported by much additional evidence (Repp, 1984:307).

It is reassuring to find that my speculations concerning the nature of the ''poetic mode'' gain support from an increasing body of experimental evidence, and that cognitive strategies have been discovered, by which listeners may switch, at will, back and forth, between phonetic categories and auditory information.

As a fourth step, one could add the vast experimental literature based on the assumption that the rich precategorial auditory information in the so-called ''encoded'' speech sounds facilitates performance in certain cognitive tasks relative to ''unencoded'' speech sounds.[2]

3. Sound symbolism and source's size

Sounds can be located along dimensions whose extremes are marked by spatial notions as LOW∼HIGH, THICK∼THIN or space-related notions as HEAVY∼LIGHT, and the like. These dimensions seem to be correlated in certain meaningful ways. There is plenty of anecdotal as well as carefully controlled experimental evidence that intuitions concerning the ''spatial'' as well as the ''tactile qualities'' of sound are fairly consistent from observer to observer, and sometimes even from culture to culture. Some such experiments have been reported by Roger Brown in his classic of psycholinguistics (Brown, 1968:110–154). The whole chapter testifies to Brown's usually brilliant insights and subtle ways of analysis. Here, however, I am going to quote only two passages with which I disagree.

> A concept like ''boulder'' is referred to rocks and stone and, in comparison, judged to be ''heavy'', ''large'', ''thick'', and ''wide''. These terms are directly applicable to boulders. However, boulders have no voices. Where, then, does the concept belong on the ''bass-treble'' or ''loud-soft'' scales? We cannot doubt the answer. If Disney were to give a boulder a voice it would be ''bass'' and ''loud'' in contrast to the piping of a pebble. This could be a mediated association: a boulder must have a bass voice because creatures that do

[2] One can discriminate by introspection that /s/ is higher than /ʃ/; but not that /da/ is higher than /ba/, and /ga/ is higher than /da/. Phoneticians call this ''encodedness'': the plosives /b/, /d/ an /g/ are more encoded than the sibilants /s/ and /ʃ/.

> have bass voices are usually heavy and boulders are heavy. It is not necessary to assume that there is any subtle inter-sensory quality found in boulders and bass voices.
>
> Subjects in the study of Brown *et al.,* felt that ''thick'' and ''thin'' simply do not apply to voices. However, ''loud'' and ''resonant'' do. Now thick people and animals and violin strings are usually loud and resonant. So, if the subject is required to guess, he will call the loud and resonant voice ''thick''. This need not be because the voice shares some inter-sensory quality with the visual or tactile apprehension of thickness. It could be because the voice is loud and creatures who have loud voices are usually thick, a mediated association (152–153).

The cognitive approach to Man, of which Brown is one of the most outstanding exponents, tends to regard such explanations as ''mediated associations'' as the last resort of the scientist, where all structural explanations fail. Now, what seems to be wrong with the ''mediated associations'' theory is that it reverts to a rather strong version of associationist theory, assuming that people in various cultures have been uniformly conditioned by external conditions. It seems to be all too easy to invent some mediating story that appears to be pretty convincing, until one becomes aware of not less convincing counter-examples. Thus, for instance, red colour is felt to be ''warm'', whereas blue is felt to be ''cold''; this feeling is not culture-dependent, and thus cannot be explained by cultural conditioning. Now there is a rather widely accepted explanation, that fire is red in all cultures whereas the blue sea is relatively cold in all cultures. However, the blue sky on a tropical (or even European) summer-noon is not exactly associated with cold. The sun, on the other hand, at its hottest, would be associated with gold rather than red, whereas red would be associated with the setting rather than with the shining sun. In this case, for instance, Brown himself offers other explanations than ''mediated association'' for ''warm'' and ''cold colours''.

I submit that bass voices are perceived as thicker than soprano voices, not because creatures that do have bass voices are usually thick and heavy, but, precisely, because ''they share some inter-sensory quality with the visual or tactile apprehension of thickness''. (I happen to know quite a few thick and heavy opera singers who have tenor or even coloratura soprano voices.) Whereas the relationship between thick people and bass voices appears to be quite incidental, the relationship between thick violin strings and ''thick'' and ''low'' sounds seems to have good physical reasons. Sounds are vibrations of the air or some other material medium. The thicker the string, *other things being equal,* the *slower* and *wider* the vibrations. (Not so with singers: when they get fatter or thinner, their voice range and voice quality remains essentially unchanged.)

There are, then, at least three physical dimensions of sound that are analogous and co-varying: SLOW∼FAST, WIDE∼NARROW, and THICK∼THIN. The first two pairs of adjectives describe the vibrations, the third pair describes the strings (if there be any) that may be causally related to the first two. It should be noted, however, that whereas the SLOW∼FAST and the WIDE∼NARROW pairs characterize the ''proximal stimulus'' that actually hits the membrane of the ear and is directly experienced, the THICK∼THIN pair characterizes the source of the sound, and may be attributed to the distal stimulus, the perceived sound, only as a concomitant of other measurable features of the soundwave (wavelength, overtone structure). Michael Polányi (1967:13) argues that the meaning of the ''proximal term of tacit knowledge'' (and, one might add, the qualities of perception) are typically displaced, away from us, to the distal term. Phenomenologically, the relative frequency and width of sound vibrations are experienced as their relative ''height'' and ''thickness'', respectively.

As for the THICK~THIN characterization of sounds, an additional observation seems to be pertinent. The sounds we usually hear do not consist of fundamentals only, but of overtones too. Since the range of frequencies audible to the human ear is limited, and since there are no ''undertones'', the lower the fundamental, the greater the number of overtones that are within the hearing range of the human ear. Thus, when we strike a key near the left end of the piano keyboard, we perceive a ''thick aura'' of overtones around the sound that is absent from the sounds produced by striking the keys near the right end (notice, by the way, that in spite of the left-to-right arrangement of the keyboard, we perceive the piano sounds as ''low'' or ''high'' rather than ''left-wing'' or ''right-wing'' as would be predicted by a mediated-association theory).

Recently I encountered a more fine-grained hypothesis which suggests a rather complex relationship between body-size, the size of articulatory organs and size of vibration, that has evolutionary implications. Even if these implications are not substantiated, the other correlations remain sound. John Ohala's paper has the telling title ''The frequency code underlies the sound-symbolic use of voice pitch''. Based on Eugene Morton's ethological work, Ohala explores some voice-pitch-related human responses, including responses to intonation. He claims that the frequency code underlying certain aspects of the sound-symbolic use of voice pitch is not merely an intercultural, but also a cross-species phenomenon. The reason is that this frequency code has great survival and evolutionary value both in mating and settling disputes:

> Animals in competition for some resource attempt to intimidate their opponent by, among other things, trying to appear as large as possible (because the larger individuals would have an advantage if, as a last resort, the matter had to be settled by actual combat). Size (or apparent size) is primarily conveyed by visual means, e.g. erecting the hair or feathers and other appendages (ears, tail feathers, wings), so that the signaler subtends a larger angle in the receiver's visual field. There are many familiar examples of this: threatening dogs erect the hair on their backs and raise their ears and tails, cats arch their backs, birds extend their wings and fan out their tail feathers. [. . .] As Morton (1977) points out, however, *the F_0 of voice can also indirectly convey an impression of the size of the signaler,* since F_0, other things being equal, is inversely related to the mass of the vibrating membrane (vocal cords in mammals, syrinx in birds), which, in turn, is correlated with overall body mass. Also, the more massive the vibrating membrane, the more likely it is that secondary vibrations could arise, thus giving rise to an irregular or ''rough'' voice quality. To give the impression of being large and dangerous, then, an antagonist should produce a vocalization as rough and as low in F_0 as possible. On the other hand, to seem small and non-threatening a vocalization which is tone-like and high in F_0 is called for. [. . .]. Morton's (1977) analysis, then, has the advantage that it provides the same motivational basis for the form of these vocalizations as had previously been given to elements of visual displays, i.e. that they convey an impression of the size of the signaler. I will henceforth call this cross-species F_0-function correlation ''the frequency code'' (Ohala, 1994:330).

Voice frequency gives, then, information not about the mass of the body, but about the mass of the vibrating membrane which, in turn, may or may not be correlated with the mass of the body. A bass singer may be slim, it is his vocal chords that must be of a substantial size.

In another paper in the same book, Eugene Morton explores avian and mammalian sounds used in hostile or ''friendly,'' appeasing contexts. He provides two tables in which sounds given by aggressive and appeasing birds and mammals are listed. ''Aggressive animals utter

low-pitched often harsh sounds, whose most general function is to increase the distance between sender and receiver. Appeasing animals use high-pitched, often tonal sounds, whose most general function is to decrease the distance or maintain close contact by reducing the fear or aggression in the receiver'' (Morton, 1994:350–353). Subsequently (353–356) he expounds a conception of sound–size symbolism in animals similar to the one quoted above from Ohala.

I am not claiming that I have enough data to confirm the evolutionary implications of the Morton–Ohala hypothesis. I am attempting to do two things: first, I am locating their hypothesis between Brown's and my own, comparing the three. Second, I claim that this hypothesis can account for certain sound–body relationships on the one hand, and certain intonation phenomena on the other that are difficult to account for otherwise. I compare three hypotheses by Roger Brown, John Ohala and myself regarding the relationship between the size of the sound source and the perceived quality of the sound. What I am showing is that when Ohala relates sound size to body size he does this, unlike Brown, via the size of the articulatory organs, corroborating (not confirming) my hypothesis. In this way he accounts also for the everyday observation that people with large body may have a thin voice, and people with a smaller body a thicker voice. Interestingly, according to Ohala and Morton, individuals in many species try to use this discrepancy between body size and voice size to fool potential partners or adversaries who act on Brown's assumptions.

As to intonation phenomena, Ohala reports a set of experiments in one of which short samples (4 s) of spontaneous speech were digitally processed in such a way as to remove all spectral details but to retain the original amplitude and F_0 contour, the latter of which was either linearly upshifted or downshifted by varying amounts or left unchanged. These samples of ''stripped speech'' were presented in pairs to listeners who were asked to judge which voice of each pair sounded more dominant or self-confident. The results indicate that, other things being equal, lower F_0 does make a voice sound more dominant. This is evident, for example, in the judgments for a pair of samples which are derived from the same speech sample but with one of them upshifted from the original by a factor of 1.25. The sample with the lower F_0 was judged as sounding more dominant than the sample with the higher F_0 by 92% of the listeners. In another experiment two samples of ''stripped speech'' were presented to listeners. A sample which had a higher-peak F_0 but ended with a sharp terminal fall was judged as sounding more dominant (92% of all judgments) than another sample that was lower in F_0 during most of its duration. ''The sharp F_0 terminal fall, lacking in the other sample, seemed to be the determining factor in listeners' evaluations; it suggests that the occasionally higher-peak F_0 in the voices exhibiting greater confidence is there in order to make the terminal fall seem to be even steeper, i.e. by virtue of having fallen from a greater height''. The Morton–Ohala hypothesis can account for such results quite plausibly. To refute it, one must present a rival hypothesis and contrive some experiment the results of which would support one or the other.

The foregoing conception may have far-reaching implications, beyond what is conspicuously suggested by Ohala and Morton. At the end of an important theoretical statement of research done at the Haskins Laboratories, Liberman (1970:321) says: ''One can reasonably expect to discover whether, in developing linguistic behavior, Nature has invented new physiological devices, or simply turned old ones to new ends''. I will suggest that in some cases at least old cognitive and physiological devices are turned to linguistic, even aesthetic, ends. This seems to reflect Nature's parsimony.

What is the relationship between being dangerous and having an irregular or ''rough'' voice quality; or between seeming non-threatening and a vocalization which is tone-like? To answer this question, one must realise that ''noises'' are irregular sounds, ''tones'' are regular,

periodic sounds. Ohala and Morton mention this merely as a corollary of ''deep'' and ''high'' voices. But this aspect of nonhuman vocalisation may throw an interesting light on certain widespread intuitions in the poetic mode of speech perception, namely, that periodic consonants (e.g., [m], [n]) are perceived as soft, mellow, and appeasing, whereas aperiodic continuants (e.g., [s], [z]) as harsh, strident, turbulent, and the like. In fact, what I wrote about the poetic effects of periodic and aperiodic speech sounds may apply, mutatis mutandis, to this ''echological'' problem as well:

> Periodic sounds have been described (May and Repp, 1982:145) as ''the recurrence of signal portions with similar structure'', whereas aperiodic stimuli as having ''randomly changing waveform'', that ''may have more idiosyncratic features to be remembered''. The recurring signal portions with similar structures may arouse in the perceiver a relatively relaxed kind of attentiveness (there will be no surprises, one may expect the same waveform to recur). Thus, periodic sounds are experienced as smoothly flowing. The randomly changing waveforms of aperiodic sounds, with their ''idiosyncratic features'', are experienced as disorder, as a disruption of the ''relaxed kind of attentiveness''. (Tsur, 1992:44)

In some circumstances unpredictability is a dangerous thing. Sound gives information about physical changes in one's environment. Randomly changing sounds give information about unpredictable changes. So they force one to be constantly on the alert. The survival purpose of such alertness is conspicuous. Even in animal communication, however, an irregular or ''rough'' voice quality is sometimes ''symbolic''; it constitutes no danger in itself, but has a common ingredient with dangerous circumstances: unpredictability. In the poetic mode of speech perception, response to regular or randomly changing waveforms is turned to an aesthetic end: it assumes ''purposiveness without purpose''. I do not offer this as a fact, or a generalisation, but as a hypothesis that would perhaps explain the relationship between unpredictability and being dangerous, by making a crucial recommendation to attend to certain aspects of the phenomenon.

Likewise, the foregoing conception may illuminate *the motor theory of speech perception* too from an unexpected angle. This theory assumes that in the production as well as in the perception of speech we attend *from* the acoustic signal *to* the combination of muscular movements that produce it (even in the case of hand-painted spectrograms); and from these elementary movements to their joint purpose, the phoneme. The best approximation to the invariance of phonemes seems to be, according to Liberman et al. (1967:43, and passim), by going back in the chain of articulatory events, beyond the shapes that underlie the locus of production, *to the commands* that produce the shapes. ''There is typically a lack of correspondence between acoustic cue and perceived phoneme, and in all these cases it appears that perception mirrors articulation more closely than sound. [. . .] This supports the assumption that the listener uses the inconstant sound as a basis for finding his way back to the articulatory gesture that produced it and thence, as it were, to the speaker's intent'' (Liberman et al., 1967:453). If Ohala and Morton are right, this mechanism underlying speech perception is a less recent invention of evolution than might be thought. The lion's roar, for instance, follows a similar course. The F_0 of voice can convey an impression of the size of the mass of the vibrating membrane and, indirectly, of the size of the signaler; in other words, the listener uses the inconstant sound as a basis for finding his way back to the articulatory organs and gestures that produced it and thence, as it were, to the roarer's intent. This does not mean that there is no qualitative leap from the lion's roar to human speech. The lion's roar can express only some general intent; not, for instance, such subtle semantic distinctions as in ''For fools *admire*, men of sense *approve*''.

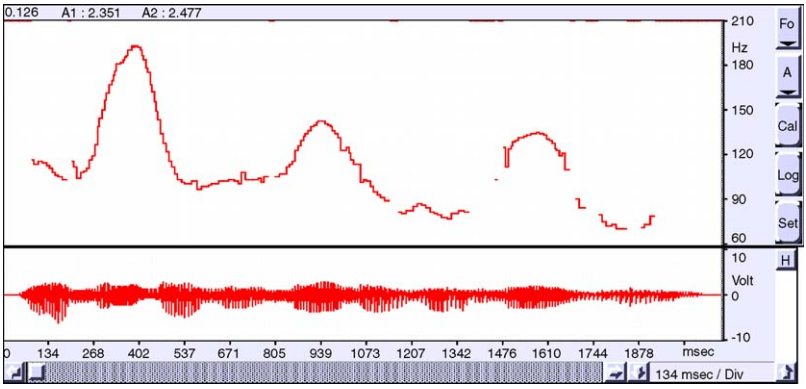

Fig. 4. Waveplot and pitch extract of an utterance in which there are no words, but in which the speaker sounds contented (after Peter Ladefoged).

After having submitted the present essay, I ran into Peter Ladefoged's illuminating book *Vowels and Consonants* with an accompanying CD, where one may listen to the sounds under discussion. He ends a discussion of the modes and moods suggested by intonation with a pair of examples (recorded in his own voice) in which ''there are no distinguishable words, but it is obvious which was spoken in anger and which when happy. In the first of these two utterances (Fig. 4) there are large changes in pitch, with the 'sentence' as a whole having a generally falling pitch. The second utterance (Fig. 5) has slightly smaller peaks, but they are sharper, without the rounded tops, and the 'sentence' as a whole has an increasing pitch'' (Ladefoged, 2001:17). (I am responsible for Figs. 1, 4–8.)

In light of Ohala's and Morton's papers, I wish to make two additional observations. First, the peaks of the second utterance are not just ''slightly smaller''; the whole utterance is of considerably lower pitch. The pitch of the highest peak in Fig. 4 is 191.739 Hz; that of the lowest peak is 135.276 Hz (my speech-analyzer application, SoundScope, specifies 80–150 as the range of typical male voice). The highest peak in Fig. 5 is slightly lower than the lowest peak in Fig. 4 (129.706 Hz); the lowest peak in Fig. 5 is 102.083 Hz. Second, the utterance reflected in Fig. 5 is uttered in a ''grating'' voice: its voice quality is quite ''rough'', considerably ''harsher''

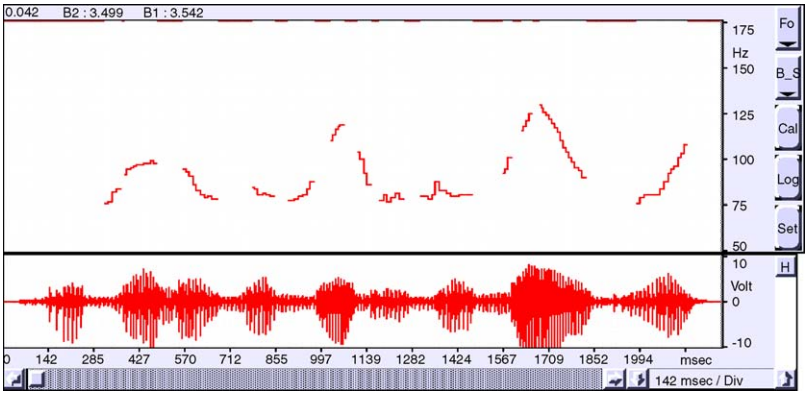

Fig. 5. Waveplot and pitch extract of an utterance in which there are no words, but in which the speaker sounds upset or angry (after Peter Ladefoged).

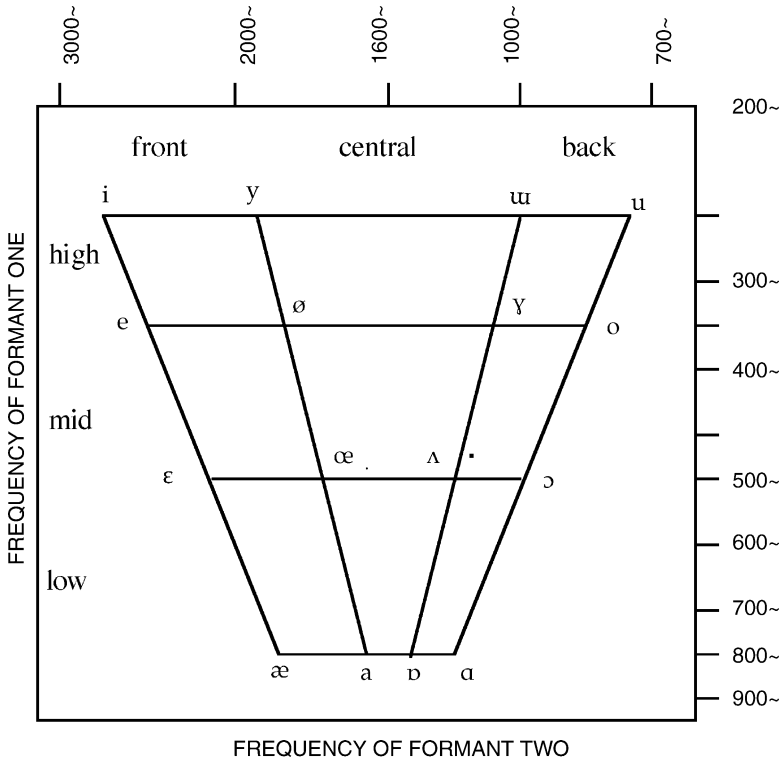

Fig. 6. The acoustic and articulatory location of the synthetic vowels, plotted according to the frequency positions of the first and second formants.

than that of the one reflected in Fig. 4. As Morton said, ‘‘aggressive animals utter low-pitched often harsh sounds’’; and so, apparently, do aggressive humans. At any rate, Ohala’s and Morton’s generalisations apply to Ladefoged’s example of upset or angry utterance too, beyond the features pointed out by Ladefoged himself. However, my instrumental analysis of

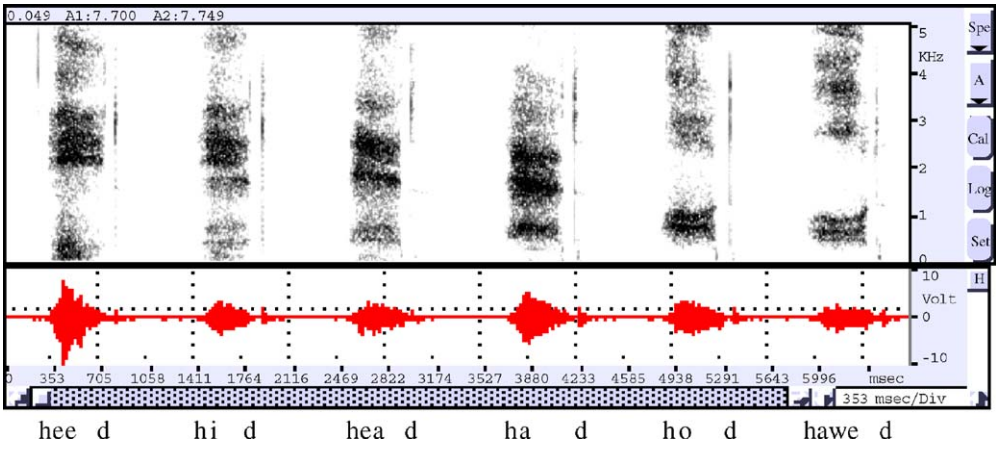

Fig. 7. Waveplot and spectrogram of the words heed, hid, head, had, hod, hawed whispered.

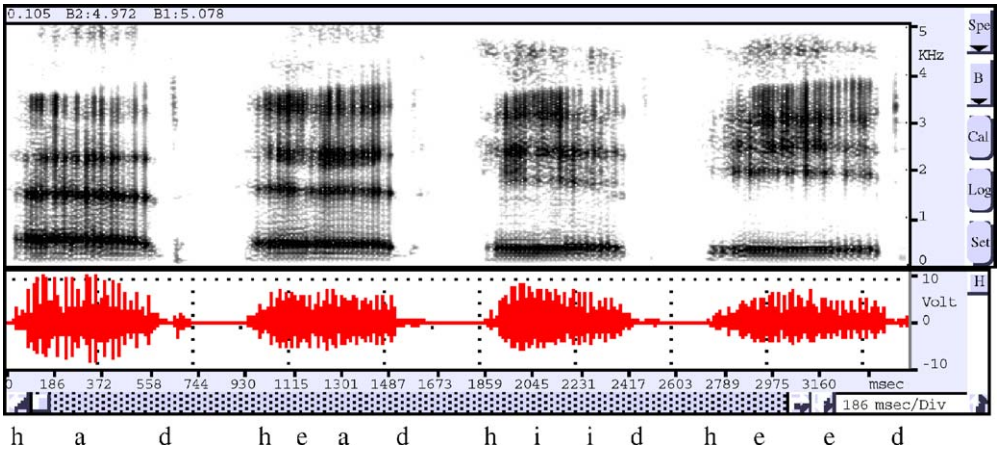

Fig. 8. Waveplot and spectrogram of the words had, head, hid, heed spoken in a creaky voice.

poetry readings suggests that only a small part of the poetic or conversational effects can be accounted for in terms of the "frequency code". High tone, for instance, may have a wide range of effects, depending on other elements with which it combines, besides submission (see Tsur, 2006a:157, 2006b).

4. Sound symbolism and referent's size

The foregoing discussion strongly suggests a causal relationship as well as structural resemblance between the frequency and perceived size of sounds on the one hand, and the physical size of their vibrating source on the other. The association of small size with high frequency and of large size with low frequency becomes a "meaning potential" of sounds, which may be actualized in sound–referent relations too. In the chapter "Some Spatial and Tactile Metaphors for Sounds" of my 1992 book I also discussed vowel symbolism for size and distance. Among others, I quoted Ultan (1978) who, by examining a total of 136 languages, tested the hypothesis that diminutive sound symbolism is associated with marked phonological features (high and/or front vowels and palatal or fronted consonants). He found that diminutive is most often symbolized by high or high front vowels, high tone, or various kinds of consonantal ablaut. Proximal distance is symbolized overwhelmingly by front or high vowels. Let me add to Ultan's sample a language not included in it, my native Hungarian, in which *itt* means "here", *ott* means "there", *ez* means "this", *az* means "that". *Így* means "in this fashion", *úgy* means "in that fashion"; *ilyen* means "of this kind", *olyan* "of that kind", and so forth. "Since high front vowels reflect proportionately higher second formant frequencies, and the higher the tone the higher the natural frequency, there appears a correspondence between a feature of high frequency (=short wavelength in physical terms) and the category of small size" (Ultan, 1978:545). Likewise, for the same reasons, the received view is that in Western languages /i/ is small and /a/ is big.[3]

[3] I wonder whether this system of front (high) vowels suggesting great distance and back (low) vowels suggesting small distance can be related to Morton's claim that aggressive animals utter low-pitched sounds, whose most general function is to increase the distance between sender and receiver, whereas appeasing animals use high-pitched sounds, whose most general function is to decrease the distance or maintain close contact by reducing the fear or aggression in the receiver.

In a mind-expanding paper on the word class of ''expressives''[4] in Bahnar, a Mon-Khmer language of Vietnam, Gérard Diffloth claims that in this word class /i/ signifies ''big'', and /a/ ''small''. This throws my foregoing argument into an exciting perspective. At first sight the paper provides outright refutation of one of my pet beliefs; but in the final resort it lends massive support to my wider conceptions, that speech events (speech sounds and articulatory gestures) do have certain (sometimes conflicting) combinational potentials, which may be activated, *after the event*, by certain meaning components. Diffloth points out the following relationships between referent size and vowel height in Bahnar[5]:

	ɨi	ɯu	i	u
''BIG''				
	ee	oo	e	o
''SMALL''	ɛɛ	ɔɔ	ɛ	ɔ

Examples (''D. red.'' = ''Descriptive reduplication''):

1. /blooŋ-blooŋ/ ''D. red. of numerous reflections caused by rays of light on a large object, elongated in shape''
   vs. /blɔɔŋ-blɔɔŋ/ ''id., small object''
2. /blooŋ-blɛɛw/ ''D. red. of the numerous reflections caused by a single ray of light on a big shiny object''
   vs. /blɔɔŋ-blɛɛw/ ''id., small shiny object''
3. /bleel-bleel/ ''D. red. of large flames appearing intermittently but remaining vivid.''
   vs. /blɛɛl-blɛɛl/ ''id., small flames''
4. /bliil-ɲip/ ''D. red. of a large scintillating fire, of the last flashes of a large fire about to die''
   vs. /blɛɛl-ƝɛP/ ''id., small fire'',

and so forth. There are examples in which a three-way gradation is given, with high vowels providing a third degree: ''enormous'':

''ENORMOUS''	ɨi	ɯu	i	u
''BIG''	ee	oo	e	o
''SMALL''	ɛɛ	ɔɔ	ɛ	ɔ

[4] ''I have used the term 'expressives' to refer to this basic part of speech, which is alien to Western tradition but can be defined in the additional way by its distinct morphology, syntactic properties, and semantic characteristics'' (Diffloth, 1994:108).

[5] Let me say at once that I know nothing about Bahnar or any other Vietnamese language except what I read in Diffloth's paper. Everything I say on this language is based on what I read in that paper.

In both the two- and three-way division ''the iconic values of the vowels are, roughly speaking: High = Big and Low = Small, exactly opposite to the English /i/ = Small and /a/ = Big, claimed to be universal. There is nothing peculiar about this Bahnar system, and one can easily find an iconic basis for it. In the articulation of high vowels, the tongue occupies a much larger volume in the mouth than it does for low vowels. The proprioceptive sensation due to this, reinforced by the amount of contact between the sides of the tongue and the upper molars, is available to all speakers and is probably necessary to achieve a precise articulatory gesture. [. . .] In this perspective, two different languages may easily use the same phonetic variable (vowel height) to convey the same range of sensations (size), and come up with exactly opposite solutions, both being equally iconic; all they need to do is focus upon different parts of the rich sensation package provided by articulatory gestures, in our case the volume of the tongue instead of the size of the air passage between it and the palate''.

Now consider such pairs of English synonyms as *big* and *large*, or *small* and *little* one member of which contains a high vowel, the other a low one. One may account for their coexistence in one language in one of two ways: either by assuming that the relationship between sound and meaning is arbitrary, or by assuming that speakers and listeners intuitively focus upon different parts of ''the rich sensation package'' provided by either the articulatory gestures or the speech signal in pronouncing these words. It may well be the case that, basically, in most words the combination of the phonetic signifier with the semantic signified is arbitrary; it is only after the event that meaning directs attention to certain aspects of the vowels in *large* or *little*, but not in *big* or *small*. There is good evidence that poets in various languages exploit this flexibility of language users: they increase the relative frequency of certain speech sounds in a poem so as to generate an emotional atmosphere, directing attention by meaning to the relevant phonetic features. The relationship between the phonetic and semantic constituents of most words used may be arbitrary; but certain features of the text's meaning may direct attention to certain recurring features of the sound patterns (Fónagy, 1961; Harshav, 1980). Shifting attention from one part of ''the rich sensation package'' to another is what Wittgenstein (1967:194 ff.) called ''aspect-switching'', prompted by the meanings of the words. According to Wittgenstein, one may switch between meanings attached to a single string of phonological signifiers; the present conception extends this ability to switching between different features of one string of phonological signifiers. In fact, there are good reasons to suppose that Wittgenstein did not mean specified visual or verbal aspects, but an ability (or lack of ability) to *switch* between aspects of whatever kinds.

There are two conspicuous common features in Diffloth's corpus and my foregoing examples from Hungarian. First, the meaning relationship, if present, does not take the shape of a statistical tendency in a huge aggregate of isolated words; it is displayed by minimal pairs of straightforward antonyms. Second, phonetically, these pairs are opposed in only one pair of vowels; semantically, too, they are contrasted in one feature. All the rest is really equal. In other words, in such cases size–sound symbolism is formally lexicalised. This lexical feature reflects creative phonetic intuitions in the distant past which have fossilised by now; the present-day language-user may attend away from the sound symbolism of ''high'' and ''low''. So, these pairs of words are structurally different from such clusters of synonyms and antonyms as *big* and *large*, or *small* and *little*. The two systems, however, are opposed in one interesting feature. In Hungarian there is vowel harmony. Consider the pair *ilyen* (ijɛn) and *olyan* (ojɔn). The size-symbolic contrast is carried by the /i∼o/ opposition; but this affects the location of the second vowel too. In Diffloth's examples from Bahnar, by contrast, the other vowels may vary independently.

I have vested theoretical interest that Diffloth's explanation should be valid. It would reinforce my conception according to which sound symbolism is part of a complex event, comprising meanings, articulatory gestures, sound waves, etc. Each one of these components has an indefinite number of features, which give rise to a multiplicity of sometimes conflicting combinational potentials. Strong intuitions concerning sound symbolism are generated by selecting a subset of available features on the semantic, acoustic, and articulatory levels. When conflicting intuitions are reported, attention is shifted from one subset to another.

When, however, I tried to pronounce the speech sounds which Diffloth designates ''high'', I noticed that his description suits [i] extremely well; but not [u].[6] In view of the examples he provides, whatever explanation suits [i] should suit [u] too.

When we compare Ultan's and Diffloth's explanations, we get a clue for solving the problem: they mean different things by the same words. Speaking of ''high'' and ''low'', Ultan means relative formant frequency; Diffloth means articulatory location.[7] Consider Fig. 6. The words ''front, central, back, high, mid, low'' refer to place of articulation. The numbers refer to formant frequency. The ''height'' of the place of articulation of the vowel is in an inverse relation to the frequency of its first formant. The higher the place of articulation, the lower is the formant frequency. In fact, we should re-write Diffloth's above statement as ''In the articulation of *front* vowels, the tongue occupies a much larger volume in the mouth than it does for *back* vowels''. This would, of course, suit the high and low vowels arranged by the frequency of the second formant, according to which /i/ is ''high'', /u/ is ''low''. But the scales of Diffloth's examples from Bahnar reflect relative frequencies of the first formant, according to which /i/ and /u/ are of equal height, /e/ and /o/ are of equal height, and so forth. So, we must assume that the conflicting sound-symbolisms of Bahnar and of Hungarian (or English) are generated not by attending to different aspects of the articulatory gesture, but by attending to different formants of the speech signal. When attending to the frequencies of the first formant, the principle of *low* is ''big'' and *high* is ''small'' is meticulously preserved in Bahnar too.

Thus, the words *high* and *low* are ambiguous in this context. If we rely on the relative height of articulation in Bahnar, *high* will be ''big'', *low* will be ''small''. If we rely on relative first-formant frequencies, high will be ''small'' and *low* will be ''big'' in Bahnar too. How can we know, then, which one is the ''correct'' identification? I have to admit that this is not clear at all. My foregoing discussion apparently provides support for both possibilities, articulatory gesture and auditory information. In proposing the ''Poetic Mode of speech perception'', I relied on ''rich precategorial auditory information''. This would favour the ''frequency code'' conception. With reference to the motor theory of speech perception, however, I quoted Liberman saying ''in all these cases it appears that perception mirrors articulation more closely than sound''. This would favour the articulatory gesture conception.[8] I propose the following way out from this muddle. By this statement, Liberman referred to the perception of phonetic categories. Perceptual and

[6] This does not imply that the much larger volume which the tongue occupies in the mouth and the larger surface of contact with the palate may not affect the perceived quality of speech sounds, e.g., their perceived wetness. Consider: ''Les consonnes palatales ou palatalisées étaient senties comme particulièrement mouillées. Par rapport á un *l* palatisée, [. . .] le *l* simple passe pour *sec*'' (Fónagy, 1979:19). Fónagy explains this judgment as follows: ''Selon les palatographies et radiographies, les occlusives amouillées, palatales ou palatalisées, se distinguent des autres par un contact nettement plus large du dos de la langue et du palais. Ceci revient a dire que la sensation kinesthésique du contact de surfaces des deux muqueuses, donc mouillées est particulièrement nette'' (Fónagy, 1979:98).

[7] Characterisitcally, Diffloth accounts for Ultan's findings, as for his own findings, in terms of articulatory gestures, not frequencies: ''in our case the volume of the tongue instead of the size of the air passage between it and the palate''.

[8] Notice, however, that the phrase ''rich precategorial auditory information'', too, is derived from Liberman.

emotional symbolism, by contrast, is founded precisely on the rich precategorial auditory information which escapes categorial perception. To be sure, articulatory gestures do have a crucial kinaesthetic effect on how speech sounds feel (see footnote 6); but we are dealing here with an auditory phenomenon: the perceived size of speech sounds. We actually perceive high-frequency sounds as higher and thinner than low-frequency sounds even when they are played on the violin or the piano, where no articulatory gestures are involved.[9]

The notion of ''consistency'' too may be relevant here. Ultan accounts for his intercultural findings with reference to second-formant frequency. Ohala speaks of the ''frequency code'' in terms of cross-species F_0-function correlation. When we apply this frequency code to first-formant frequencies in Diffloth's findings in Bahnar, they become consistent with earlier findings in other languages. ''The records show that there are well-developed sound-symbolic systems where vowel quality is used with systematic results exactly opposite to those predicted'' (Diffloth, 1994:107)—provided that we change the rules of the game. I know, of course, nothing about the phonetic intuitions of Bahnar-speakers, or their pronunciation; nor did Diffloth make any claims about them.

Having said all this, I had some doubts about my own argument. What if the second formant is simply more salient in the stream of speech, and we have no (even subliminal) access to the first formant? If this is so, all my speculations are refuted. Indeed, as Ladefoged (2001:64) suggested (and demonstrated) in a context of speech synthesis, ''the second formant by itself conveys more information than the first''. But he also demonstrates (Ladefoged, 2001:33) that by certain vocal manipulations in natural speech we *can* direct attention, even at will, to one formant or another. By whispering or using a low, creaky voice one may tune in to different formants: ''Try whispering *heed*, *hid*, *head*, *had*, *hod*, *hawed* [...]; there will be a general impression of a descending pitch''. The same sequence of vowels, however, may be arranged in a reverse order, and still heard as descending, if uttered in a low, creaky voice: ''When saying the words *had*, *head*, *hid*, *heed* [in a creaky voice], this pitch goes down''. The reason for this difference is that in these two conditions we are tuned to vibrations of the air in different parts of the vocal tract. ''The sound that you hear when whispering is mainly that of the vibrations of the air in the front of the mouth. Conversely, the pitch changes associated with saying *had*, *head*, *hid*, *heed* in a creaky voice are due to the vibrations of the air in the back of the vocal tract. This resonance is the lower in pitch of the two, and is called the first formant'' (listen online to Ladefoged uttering these sequences). I don't mean to suggest that Bahnar speakers resort to a creaky voice in order to attend to the first formant, just as we, in our Western culture, don't resort to whispering in order to perceive the second formant on which our size–sound symbolism is based. Perhaps there are less palpable techniques to shift attention between the front and the back of the vocal tract. But Ladefoged's exercises prove that switching attention between first and second formants *is possible* at least. There may be good reasons for attending to the second rather than the first formant. But this still does not rule out the possibility that in some culture in general, or regarding one particular ''expressive'' part of speech, speakers attend to the first formant. What is more, there are legitimate linguistic and paralinguistic contexts in the world's languages in which creaky voice would be natural in one way or other, and might elicit sound-symbolic intuitions.

[9] To avoid misunderstanding: I am not trying to explain how two different modes of perception can work to justify both Ultan and Diffloth. I am trying to explain how two different modes of perception (the speech mode and the nonspeech mode) can eliminate, within Liberman's and my conception, the contradiction I have pointed out between two opposite claims: claiming that articulatory gestures rather than precategorial auditory information determine perception and claiming its converse.

The following thought experiment would be consistent with the possibility that speakers can shift attention to the first or the second formant at will. Wittgenstein (1967:194 ff.) says that ''aspect-switching'' is like understanding the request to pronounce the word ''till'' and mean it as a verb, or an adverb; or the word ''March'' and mean it once as an imperative verb, and once as the name of a month. ''You can say the word 'March' to yourself and mean it at one time as an imperative at another as the name of a month. And now say 'March'!—and then 'March *no further*!'—Does the *same* experience accompany the word both times—are you sure?'' (ibid., 215). Now suppose I strike a note near the left end of the piano keyboard, then a note near the right end. You will hear the right-end note as ''higher'' than the left-end note. Suppose, however, that I request you to experience them as equally high, or once the right note, once the left note as higher. You will say that you don't know what I am talking about. Now suppose I ask you to pronounce the vowels /i/ and /u/ on the same pitch and experience /i/ as higher. You will do this without any difficulty. If I ask you to experience them as equally high, you can do this too and experience the switch in a way that is similar to the switching between the meanings of ''till'' or ''March''. If, however, I ask you to experience /u/ as higher than /i/, you again will say that you don't know what I am talking about. A glance at Fig. 2 may easily account for this. When you experience /i/ as higher, you attend to the second formant; when you experience the two vowels as equally high, you attend to the first formant.

When I first read Diffloth's paper, I thought that his examples were counterexamples to the widespread belief (which I too entertained), that high sounds (including high vowels) tended to suggest small referents, whereas low sounds and vowels large referents. By the same token, I thought, it supported my broader generalisations concerning human flexibility in switching between various aspects of the same speech sounds. His explanation, however, based on articulatory gestures, conflicted with the linguistic facts he adduced. The present paper proposes an analysis that elucidated the problem and lends support to both of my former beliefs. I am most sympathetic with Diffloth in ''deploring the incorrect use of the term 'universal' to mean simply 'found in a number of languages'''. But, as far as the present issue is concerned, the convincing counterexamples are still to be adduced.

5. Methodological comments

Some of my arguments in this paper are not exactly of the kind linguists would expect and can, therefore, easily be misunderstood. So, in order to avoid misunderstandings and to elucidate my point, I propose to make a few methodological comments. On my discussion of Bahnar, for instance, the referee quoted above on phlogiston said that I was relying on one very obscure language and a single case study of certain aspects of it. ''The conclusions from this thin material are stretched too thin for the universal validity that Tsur wants to find''. This arouses an intriguing methodological problem. The validity of my argument depends here, to a considerable extent, on the purpose of my discussion. Had I tried to find some universal validity, that criticism would have certainly been fatal to my argument. But I was engaged in a very different kind of activity: I was trying to expose a hidden inconsistency in Diffloth's argument. I was not universalising, I was refuting an alleged counterexample to a received view by comparing the rules of the game followed by two scholars, as well as the outcomes of their inquiries. Pointing out that two scholars obtain different results by following different games and that one obtains similar results if one follows the rules of the same game with reference to the two corpora in question does not need a huge database. It requires just these two arguments to compare.

Now consider the following issue. The afore-mentioned referee said that in my discussion of section 3 my counterexamples to Brown were not more persuasive than his; but he did not explain why. Elsewhere he objects to certain leaps in my argument. So, I can do little better than make explicit some of my implicit assumptions. Physicists have unambiguously established that the size of strings co-varies with their speed of vibration, and that speed of vibration co-varies with wavelength (that is, with the size of the wave). This is certainly quite rigorous. From the direction of psychology, there is an ubiquitous perception that when one strikes a note near the left end of the piano keyboard (where strings are thicker) it sounds deeper and thicker than a note at the right end. At this point a leap becomes inevitable. When I speak of the causal relationship between the measurable thickness of a string and the perceived thickness of the sound produced by it, I am committing an unavoidable leap. I must *assume* that the perception of thickness is mediated by the measurable length of the sound waves produced and the measurable amount of overtones in the perceptible range that strike the perceiver's ear membrane and outer skin. More precisely, I must assume that these features of the wave are perceived as ''thickness'', offering this as a *plausible hypothesis* to account for the similarity of the perceived sound and its source. I am not claiming that there is hundred percent certainty that this is so, only that this hypothesis is more *plausible* than Brown's.

No amount of measurement can prove or disprove a causal relationship between the measurable thickness of a string and the perceived thickness of the sound produced by it. In such leaps, in J.J.C. Smart's (1966) words in a context of theory construction in physics, ''expressions like 'make more plausible', 'lead us to expect that', or 'strongly suggest' apply, but where the logical relations of implication and contradiction do not strictly apply'' (239). Smart points out that '''rigour', in the sense in which it is pursued in pure mathematics is not an ideal in applied mathematics (or physics). The conception of 'rigour' involved in physics is that whereby it makes sense to say 'rigorous enough'''' (ibid., 237). It would not be too far-fetched to claim that in literary theory, speech perception and other human disciplines less rigour is 'enough' for leaps even than in physics. But, in any case, if you want to do measurements, you must start out with a hypothesis, which I provide. The experimentalist will have to contrive an experiment to decide whether the perceived thickness of sound is correlated with the mass of one's body or with that of one's vibrating membrane.

References

Brown, Roger, 1968. Words and Things. The Free Press, New York.

Diffloth, Gérard, 1994. i: *big*, a: small. In: Hinton, Leanne, Nichols, Johanna, Ohala, John J. (Eds.), Sound Symbolism. Cambridge University Press, Cambridge, pp. 107–114.

Fónagy, Iván, 1961. Communication in poetry. Word 17, 194–218.

Fónagy, Iván, 1979. La Métaphore en Phonétique. Didier, Ottawa.

Harshav, Benjamin, 1980. The meaning of sound patterns in poetry: an interaction view. Poetics Today 2, 39–56.

Hrushovski, Benjamin, 1968. Do sounds have meaning? The problem of expressiveness of sound-patterns in poetry. Hasifrut 1, 410–420 (in Hebrew). English Summary: 444.

Ladefoged, Peter, 2001. Vowels and Consonants—An Introduction to the Sounds of Languages. Blackwell, Oxford.

Liberman, A.M., 1970. The grammars of speech and language. Cognitive Psychology 1, 301–323.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. Psychological Review 74, 431–461.

May, Janet, Repp, Bruno H., 1982. Periodicity and auditory memory. In: Status Report on Speech Research SR-69, Haskins Laboratories, pp. 145–149.

Morton, Eugene S., 1994. Sound symbolism and its role in non-human vertebrate communication. In: Hinton, Leanne, Nichols, Johanna, Ohala, John J. (Eds.), Sound Symbolism. Cambridge University Press, Cambridge, pp. 348–365.

Ohala, John J., 1994. The frequency code underlies the sound-symbolic use of voice pitch. In: Hinton, Leanne, Nichols, Johanna, Ohala, John J. (Eds.), Sound Symbolism. Cambridge University Press, Cambridge, pp. 325–347.

Polányi, Michael, 1967. The Tacit Dimension. Anchor Books, Garden City, NY.

Repp, Bruno H., 1984. Categorical perception: Issues, methods, findings. In: Lass, N.J. (Ed.), Speech and Language: Advances in Basic Research and Practice, vol. 10. Academic Press, New York, pp. 243–335.

Smart, J.J.C., 1966. Theory construction. In: Flew, A.G.N. (Ed.), Logic and Language. Blackwell, Oxford, pp. 222–242.

Tsur, Reuven, 1992. What Makes Sound Patterns Expressive: The Poetic Mode of Speech-Perception. Duke University Press, Durham, NC.

Tsur, Reuven, 2001. Onomatopoeia: cuckoo-language and tick-tocking—the constraints of semiotic systems. In: Iconicity In Language. Available online: http://www.trismegistos.com/IconicityInLanguage/Articles/Tsur/default.html.

Tsur, Reuven, 2006a. ''Kubla Khan''—Poetic Structure, Hypnotic Quality and Cognitive Style: A study in mental, vocal, and critical performance. John Benjamins, Amsterdam.

Tsur, Reuven, 2006b. Delivery style and listener response in the rhythmical performance of Shakespeare's sonnets. College Literature 33 (1), 170–196.

Ultan, Russell, 1978. Size-sound symbolism. In: Greenberg, Joseph H. (Ed.), Universals of Human Language, Volume 2: Phonology. Stanford University Press, Stanford.

Wittgenstein, Ludwig, 1967. Philosopical Investigations (translated by G.E.M Anscombe), Blackwell, Oxford.

Reuven Tsur is Professor emeritus of Hebrew Literature at Tel Aviv University, and Middle East vice president of the International Association of Empirical Aesthetics. He has developed a theory of Cognitive Poetics, and applied it to rhyme, sound symbolism, poetic rhythm, metaphor, poetry and altered states of consciousness, period style, genre, archetypal patterns, translation theory, and critical activities. His books in English include *"Kubla Khan"—Poetic Structure, Hypnotic Quality and Cognitive Style: A Study in Mental, Vocal, and Critical Performance* (2006), *On the Shore of Nothingness—A Study in Cognitive Poetics* (2003), *Poetic Rhythm: Structure and Performance—An Empirical Study in Cognitive Poetics* (1998), *Toward a Theory of Cognitive Poetics* (1992), *What Makes Sound Patterns Expressive: The Poetic Mode of Speech-Perception* (1992), *On Metaphoring* (1987), *A Perception-Oriented Theory of Metre* (1977). His non-academic publications include volumes of poetry translation into Hebrew, and memoirs from the Holocaust (in Hebrew).