Tsur, Reuven. What Makes Sound Patterns Expressive? Durham, NC: Duke UP, 1992. 1-9, 20-35.

How Do Sound Patterns Know They Are Expressive? The Poetic Mode of Speech Perception

Expressive Sound Patterns

Children often ask: "How does the dog know that barking dogs don't bite"? Similarly, it seems worth asking: May we attribute to dogs and sounds just any property we like, or is there something in the *nature* of dogs and sounds that warrants the attribution of these properties and renders their behavior consistent? Literary critics and ordinary readers usually have strong intuitions about the expressiveness of sound patterns in poetry. A vast literature exists on the subject; however, much of it is ad hoc, arbitrary, or skeptical. "It is precisely critics interested in the meaning and idea-content of poetry," says Hrushovski (1968: 410), "that feel some kind of embarrassment toward the existence of sound organization, and attempt to enlist it in the service of the total interpretation. As against this approach there are critics and theoreticians who deny all in all the very existence of specific meanings attributable to specific sounds."

In what follows I shall adopt Hrushovski's approach, according to which the various language sounds have certain general potentialities of meaningful impression (412) and can be combined with other elements so that they impress the reader as if they expressed some specific meaning (411). My claim is that these general potentialities--which I shall refer to as combinational potential--have firm, intersubjective foundations on the acoustic, phonetic, or phonological levels of the sound structure of language. More specifically, I shall rely on a simplified version of the mechanism as put forward by Liberman and his colleagues at the Haskins Laboratories, by which listeners decode the sounds and recover the phonemes.

Hrushovski claims that much of the dispute over whether sound can or cannot be expressive comes to a dead end because the issue is treated as if it were one phenomenon. "As a matter of fact, there are several kinds of relations between sound and meaning, and in each kind the problem is revealed in different forms" (412). He discusses four kinds of such relations: (a) Onomatopoeia; (b) Expressive Sounds; (c) Focusing Sound Patterns; (d) Neutral Sound Patterns. The main business of this book concerns the second of these relations, but my discussions will have some implications for the first. Hrushovski describes expressive sound pattern as follows: "A sound combination is grasped as expressive of the tone, mood or some general quality of meaning. Here, an abstraction from the sound pattern (i.e. some kind of tone or 'quality' of the sounds is parallel to an abstraction from the meaning of the words (tone, mood etc.)" (444).

Traditional poetics has important things to say about how "tone, mood etc." are abstracted from the meaning of the words. But how are they abstracted from the speech sounds? In this chapter I shall look into some possible sources of the "tone" or "quality" of the sounds and the way that tone or quality is grasped in relation to an abstraction from the *meaning* of the words (tone, mood, emotion, etc.). One important aspect of the issue is that sounds are what I call "double-edged"; that is, they may be expressive of vastly different, or even opposing, qualities. Thus, the sibilants /s/ and /š/ may have a *hushing* quality in one context and a *harsh* quality to varying degrees in some others. Hrushovski quotes Poe's line

And the silken, sad, uncertain rustling of each purple curtain

where the sibilants may be onomatopoetic, imitating the noises; or they may reinforce--or be expressive of--a quiet mood in Shakespeare's sonnet:

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.

My argument relies on the assumption that sounds are bundles of features on the acoustic, phonetic, and phonological levels. The various features may have different expressive potentialities. The claim I shall elaborate is that in different contexts, different potentialities of the various features of the same sounds may be realized. Thus, the sibilants /s/ and /š/ at *some* level of description may have features with noisy potential and others with hushing potential. In Poe's line the former is realized by the contents, in Shakespeare's quatrain the latter.

At the beginning of Fónagy's article on communication in poetry (1961), statistical methods are applied to the expressive correspondence between mood and sound quality in poetry. This work is of particular interest for at least two reasons. First, it does not investigate the relations of sounds with specific themes, but with highly generic moods: tender and aggressive. Second, it does not consider these moods in isolation, but as a pair of opposites whose mutual relations may be treated in terms of more/less rather than in absolute terms. The data Fónagy presents are illuminating and highly suggestive. In six especially tender and six especially aggressive poems by the Hungarian poet *Sándor Petőfi*,

the majority of sounds occur with the same relative frequency in both groups. All the more striking is the fact that the frequency of certain sounds shows a significant difference in both groups. The phonemes /1/, /m/, and /n/ are definitely more frequent in tendertoned poems, whereas /k/, /t/, and /r/ predominate in those with aggressive tone. For some reason, precisely these sounds seem to be the most significantly correlated with aggression, either positively, or negatively. (195)

The phonemes /m/ and /n/ have a similar negative correlation with aggression in poems by Hugo and Verlaine; /l/ is overwhelmingly tender for Verlaine, but not for Hugo. The voiceless stops /k/ and /t/ are significantly less frequent in tender poems by Petőfi, Verlaine and Hugo, and Rückert (Hungarian, French, and German poets). So, this distribution is surely not language-dependent. It would be interesting to know, to what extent if at all, "double-edgedness" is responsible for the equal distribution of other sounds in both groups of poems, owing to conflicting features' canceling out each others' influence (I shall try to answer this question later on). As for vowels, Fónagy mentions Macdermott who, through a statistical analysis of English poems, found that dark vowels are more frequent in lines referring to dark colors, mystic obscurity, or slow and heavy movement, or depicting hatred and struggle (Fónagy, 1961: 194). From this summary, one might expect to find a greater frequency of dark vowels in aggressive poems than in tender ones. Fónagy's investigation of Petőfi's poetry reveals that this is indeed the case (for the other poets, he gives only the consonant distribution). Whereas dark vowels occurred in Standard Hungarian 38.88 percent of the time, in Petőfi's aggressive poems it was 44.38 percent, and in his tender poems 36.73 percent. We receive a reverse picture from the distribution of light vowels. In Standard Hungarian they occur 60.92 percent of the time, whereas in the tender poems it was 63.27 percent, and 55.62 percent in the aggressive poems. While these deviations from Standard Hungarian seem to be convincing enough, one might reasonably conjecture that the correlation between aggressive mood and dark vowels may be even more compelling. The point is that the results may have been "contaminated." A poem may have an especially tender mood and still refer to dark colors (which, in turn, would have induced the poet to use words with dark vowels). The list of tender poems examined by Fónagy suggests that this may be the case in his corpus. Two of the tender poems seem to have dark atmospheres (or themes, at least): *"Borús, ködös őszi idő" ("Dark and Foggy Autumn Weather")*, and *"Alkony" ("Dusk")*. Statistical methods in poetics do not seem to be very successful in handling such multidimensional contrasts and correlations between moods and qualities.

Recent structuralist techniques make it possible to contrast several dimensions simultaneously; these, in turn, may bring up a considerable number of meaning and sound components, which may combine in a variety of ways. Let us consider such a "minimal pair," thought up by Richards (1929: 220) for a somewhat different purpose. One of the many sacred cows he cheerfully slaughters in *Practical Criticism* is "the notion that poetic rhythm is independent of sense."

It is easy, however, to show how much the rhythm we *ascribe* to words (and even their inherent rhythm's sounds) is influenced by our apprehension of their meanings. Compare, for example:--

Deep into a gloomy grot with

Peep into a roomy cot.

"Gloomy grot" and "roomy cot" are contrasted by, roughly, such semantic features as CONFINED□SPACIOUS; ILL-LIGHTED□BROAD DAYLIGHT; DISMAL□LIGHTSOME; SUBTERRANEAN□ON-THE-SURFACE; UNEARTHLY□EARTHLY; GRAVE□EVERYDAY; GRAVE□LIGHT. *Deep* and *peep* are contrasted by such semantic features as (FAR) DOWNWARD□UPWARD ("TO PEEP OVER"); GRAVE□FURTIVE; HEAVY□NIMBLE. Some of these contrasting pairs affect the rhythmic movement of these phrases (via, perhaps, our performance), resulting in a heavy, slow cadence in the former and a light rhythm in the latter (the heavy utterance of the former also uses the consonant clusters /gl/ and /gr/ where in the latter there are nonalliterative single consonants). However, in this case performance only reinforces a feeling of heaviness or lightness generated by these features. But, owing largely to the act of contrasting, one also becomes aware in the back of one's mind of some interaction between semantic and phonetic features. Consider the stressed long vowels shared by the contrasted words:

-eep	into a	**oomy**	**-ot**
long		long	short
high		high	middle
bright		dark	dark

In each of the two phrases different vowel features may be used to enhance meaning; this is the source of the double-edgedness of the sounds. In *peep* one tends to foreground the features [BRIGHT, HIGH], in *deep* the features [LONG, (FAR) DOWN]. In *gloomy* the feature [DARK] whereas in *roomy* the features [LONG, HIGH] (that is, spacious) are likely to be foregrounded. One is, indeed, tempted to quote Pope outrageously out of context, that is, with an emphasis on *seem*:

The sound must *seem* an echo to the sense.

Sound Color

The phrase "vowel color" can be used in three different senses. The most obvious one implies what is usually referred to as *audition colorée*, in which each vowel is consistently associated with a specific color in the consciousness of certain people (for an illuminating account of a rare case of such colored hearing, see Reichard et al., 1949; see also chapter 4). The second sense refers to an association of certain oppositions of groups of vowels with certain oppositions of abstract properties of colors. Thus, the opposition FRONT VOWELS□BACK VOWELS is associated with the opposition BRIGHT□DARK; and the opposition LOW□HIGH vowels is perceived as CHROMATIC□ACHROMATIC, and is associated with MORE□LESS VARIEGATED colors. These associations of oppositions seem to have considerable intersubjective and intercultural validity. Vowel colors in these two senses are related in the way "specific" and "general" are related. "The unambiguous tendency to feel that back vowels are 'darker' and the front vowels are 'lighter' finds further support in the assignment of darker colors to back vowels and light colors to front vowels by diverse kinds of observers" (Jakobson and Waugh, 1979: 188). At least one extremely important article (Delattre et al., 1952) uses the phrase in a third sense: the distinctive quality of each vowel as it appears to consciousness.

The first sense refers to a phenomenon whose use for poetics is not quite clear, though very interesting from the psychological point of view and fairly consistent from informant to informant (with occasional deviations). Though the notion *audition colorée* is usually invoked in discussions of Rimbaud *"Voyelles,"* the sonnet does not obey its rules. It associates the color "red" with the vowel /i/, for example, whereas in "genuine colored hearing" it is usually associated with /a/. We ought to look, then, for poetic significance on more abstract, less specific levels (see chapter 4). Later I shall explore the possible intersubjective basis of the association of the opposition FRONT□BACK vowels with the opposition BRIGHTNESS□DARKNESS as well as the possible relationship of the formant structure of particular sounds with other tone color qualities. During the past thirty years or so there has been an enormous breakthrough in our understanding the relationship between perceived speech sounds and the acoustic signal that carries them, but very little of this has reached literary theory and criticism. I am going to draw on some of this knowledge.

As a first approximation, I wish to point out that tone color refers, in general, to a property of sounds, the ecological value of which suggests that it may have preceded the development of language. It refers to that characteristic quality of sound, independent of pitch and loudness, from which its source or manner of production can be inferred; the quality of sound from which we infer, for instance, that what has fallen is a piece of wood or a piece of metal. The color of a sound is determined by its overtone structure. Overtones are sounds, higher in frequency than the fundamental, simultaneously emitted with it.

But instead of this chord which should often sound quite agreeable, we usually hear a single tone, the fundamental. The others are "repressed" and replaced by the experience of tone color which is "projected" onto the audible fundamental. . . . Without tone color fusion we would have to analyze the complex and often confusingly similar composition of the overtone chords, in order to infer the substance of the sounding things and identify them. Hence, a conscious overtone perception, if it were at all possible, would be biologically less serviceable. (Ehrenzweig, 1965: 154)

I wish to make three comments on this description. First, the perception of overtones is impossible, in many cases, owing to physiological limitations of the human ear. The fine discriminations it would require exceeds the ear's capacity, and so one is able to get only a general impression of the overtone structure of the sound. [1] Second, as we shall see in the discussion by Delattre et al. (1952), "tone color fusion" is not a unitary process and may involve several degrees and types of fusion. Third, Ehrenzweig's Freudian terminology of "repressing" and "projecting" ought to be supplemented by some other terminology--for example, Polányi's.

In Polányi's terms (1967: 10 - 11), we might say that we have an instance of tacit knowledge, that is, of knowing more than we can tell. We know the difference between the click of a metallic object and that of a wooden object, but we cannot tell how we know this. We *attend from* the proximal term, the overtone structure of the sound, *to* the distal term, its tone color; just as in the case of human physiognomy we are attending *from* our awareness of its features *to* the characteristic appearance of a face and thus may be unable to specify the features; or as we are attending *from* a combination of our muscular acts to the performance of a skill. "We are attending *from* these elementary movements *to* the achievement of their purpose, and hence are unable to specify these elementary acts. We may call this the *functional structure* of tacit know-

[1] The limitations of the human ear seem to have had beneficial effects on the organism's achievement of functional solutions. In listening to speech, people can perceive as many as 25 or 30 phonetic segments per second; this rate would far overreach the temporal resolving power of the ear. "Discrete acoustic events at that rate would merge into an unanalyzable buzz," though "a listener might be able to tell from the pitch of the buzz how fast the speaker was talking" (Liberman et al., 1967: 432). "But given that the message segments are, in fact, encoded into acoustic segments of roughly syllabic size, the limit is set not by the number of phonetic segments per unit time but by the number of syllables. This represents a considerable gain in the rate at which message segments can be perceived" (Liberman et al., 1972). This elegant solution is the result of what we shall call *parallel transmission*.

ing." Moreover, we may say that "we are aware of the proximal term of an act of tacit knowing in the appearance of its distal term." In the case of tone color, we may say that we are aware of the overtone structure of a sound in terms of the metallic click to which we are attending from it, just as we are aware of the individual features of a human physiognomy in terms of its appearance to which we are attending (or as we are aware of the several muscular moves in the exercise of a skill in the performance to which our attention is directed). This we may call the *phenomenal structure* of tacit knowing. As for the *semantic* aspect of tacit knowing, let me quote only Polányi's concluding remark on this subject: "All meaning tends to be displaced *away from ourselves*, and that is, indeed, my justification for using the terms 'proximal' and 'distal' to describe the first and second terms of tacit knowing" (ibid., 13).

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." The present suggestion is twofold. In some cases, at least, cognitive and physiological devices are turned to linguistic ends. This seems to reflect nature's parsimony. It is by now well established that acoustic signals for vowel perception are overtones; whereas the fundamental frequency may vary with the pitch of the speaker's voice, the vowel formant frequencies vary mainly (but not exclusively) with vowel color (in the third sense). Overtones are substantial ingredients in voiced consonants too.

Acoustic Coding

As we shall see, there is no one-to-one relationship between the segments of perceived speech and the segments of the acoustic signal that carries it. According to the Haskins theory of speech perception (for example, Liberman, 1970; Liberman et al., 1967), there is between the two a mediating step of "complex coding." The same coding can be "cued" by different acoustic signals, whereas sometimes the same acoustic signal may cue different phonemes. I have suggested that an important way by which this may be achieved can easily be described in the terms I have put forward. In production as well as in perception, we attend *from* the acoustic signal *to* the combination of muscular movements that produce it; and from these elementary movements, we attend 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. At the perception end, this is frequently called the *motor theory of speech perception*.

Some poetic implications of this conception seem to be as follows: our tacit knowing of the acoustic-linguistic message, in the course of perception, seems to proceed by attending *from* the acoustic signal *to* the combination of muscular movements and, further away, to their joint purpose, the phoneme. Nonetheless, in certain circumstances, which we might call the "poetic mode," some aspects of the formant structure of the acoustic signal may vaguely enter consciousness. As a result, people may have intuitions that certain vowel contrasts correspond to BRIGHTNESS□DARKNESS, or that certain consonants sound more metallic than others. As a result, poets may more frequently use words that contain dark vowels in lines referring to dark colors, mystic obscurity, or slow and heavy movement, or in depicting hatred and struggle. At the reception end, readers may have vague intuitions that the sound patterns of these lines are somehow expressive of their atmosphere.

To spell out in detail the foregoing generalizations: Vowels consist of specific combinations of overtones, called formants. A formant is a concentration of acoustic energy within a restricted frequency region. These concentrations of energy can be converted into patches of light and shade called spectrograms, with the help of a device called a spectrograph. Three or four formants are usually seen in spectrograms of speech. In the synthetic, hand-painted spectrograms of figures 1 and 2, only the lowest two are represented. Formants are referred to by numbers as F_1, F_2, etc., the first being the lowest in frequency, the next the next higher, and so on (F_0 refers to the fundamental pitch). A formant transition is a relatively rapid change in the position of the formant on the frequency scale. A device called pattern-playback converts hand-painted spectrograms into sound. This provided the basis for what proved to be a convenient method of experimenting with the speech signal: it made it possible to vary those parameters that were guessed to be of linguistic importance and then hear the effect.

Each pattern of Figure 1 consists of two bands of acoustic energy called "formants." At the left, or beginning of each pattern the formants move rapidly through a range of frequencies. The rapid movements, which consume about 50 msec, are called "transi-

Thought-Experiment with "Dark" Vowels

Let us turn to a second thought-experiment, based on another set of real experiments at the Haskins Laboratories back in the early fifties, that offers an explanation for our intuition that back vowels are darker than front vowels. Certain physical qualities of the acoustic signal enter consciousness, in spite of the speech mode, when we perceive back vowels as dark and front vowels as bright. These associations seem to have general, culture-independent validity. The search for an explanation is summarized by Jakobson and Waugh (1979: 188-94).

I submit, in accordance with the structuralist conception, that it is more adequate to conceive of the back-front continuum as analogous in some way to the dark-light continuum than to attribute specific properties to the individual vowels. To this effect, Jakobson and Waugh (ibid., 189) quote Gombrich with approval:

The problem of synaesthetic equivalences will cease to look embarrassingly arbitrary if, hereto, we fix our attention not on likeness of elements but on structural relationships within a scale or a matrix. When we say that *i* is brighter than *u*, we find a surprising degree of general consent. If we are more careful still and say that the step from *u* to *i* is more like an upward step than a downward step, I think the majority will agree, whatever explanation each of us may be inclined to offer.

In other words, there need not be anything inherently dark in the vowel /u/, or inherently bright in the vowel /i/. Suffice that the vowel continuum and the brightness continuum are perceived as somehow analogous to one another, and that the vowel /u/ and the value dark occupy similar positions on them. However, we still have to explain why the bright end of the brightness continuum is matched precisely with the front extreme of the vowel continuum and not the other way around; or, likewise, why the high end of the height continuum is matched with the /i/ end of the vowel continuum rather than the other way around. My claim is, again, that certain perceptual aspects of the acoustic signal, irrelevant in principle to the speech mode, do enter consciousness in spite of all. Consider the widespread intuition expressed by Gombrich that the step from /u/ to /i/ is more upward than downward. In an important sense, each vowel can be uttered on any fundamental pitch. Yet in another, no less important, sense the two vowels are exactly the same height. This will be apparent if we look again at figure 2. The first formants of the two vowels are of exactly the same frequency (250 cps). If, on the other hand, we look at the frequency of the second formants, we find that the one for /i/ is of a much higher frequency than the one for /u/ (2,900 vs. 700 cps; for the numbers, see Delattre et al., 1952: 198). It should be noted that this intuition of an upward step from /u/ to /i/ can easily be overridden when the fundamental pitch pattern goes in the opposite direction, as in *intrude*!

Now, what about the correspondence of the dark-bright continuum to the /u/-/i/ continuum? Let us grant that these continua should be made analogous. Why should the /i/ pole be matched precisely with the bright pole and not the other way around? In what acoustic features relevant to the opposition DARK□BRIGHT are the back vowels opposed to the front vowels? Looking at figure 2, we might say that the significantly higher second formant bestows some luster on the front vowels. This assumption is corroborated by the effects of reducing the intensity of the higher formant of most front vowels and some middle vowels. "Small attenuations of the higher formant caused the vowel to acquire a quality that can best be described as 'dull'" (Delattre et al., 1952: 204).

But a much more interesting side to the story emerges, having considerable psychological consequences. In figure 2 the three rightmost spectrograms are of back vowels, the three leftmost of front vowels. In the former, the two formants are significantly closer to one another than in the latter. Figure 4 shows, though in a less transparent manner, that the same kind of correlation is true of all sixteen cardinal vowels of the IPA. Delattre and his colleagues produced good experimental evidence that vowels whose first and second formants are closer together are perceived differently from those whose formants are wider apart. Roughly, the human ear effectively fuses the two formants when they are close enough to one another, whereas it seems to perceive them as fairly differentiated when they are sufficiently apart. In an important sense, the formant structure of the front vowels is more clearly articulated. Alternatively, one might say that the ear averages the first and second formant frequencies of the back vowels, whereas the first two formants of the front vowels, though perceived categorically, clearly stand out in our subliminal perception. For the vowels where the first

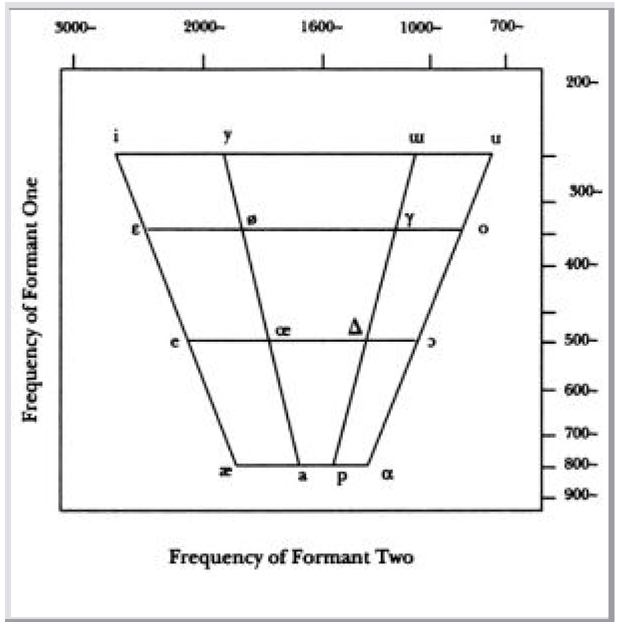

Figure 4.
The acoustic location of the synthetic vowels, plotted according to the frequency positions of the first and second formants.

two formants are relatively close together, one can find reasonably good one-formant equivalents that occupy an intermediate position between the two-formant vowels (nearer to the lower formant). It is difficult to find one-formant equivalents for the front vowels in which the two formants are rather far apart, except for /i/, which has the greatest separation of the first and second formants; here, a single formant near the normal position of the *second* formant seems to produce the /i/ color rather well. "/i/ is an extreme vowel (none has a higher second formant), and the kind of judgment-anchoring which so often occurs at the extremes of a stimulus series might account for the fact that a high-frequency formant was so often identified as /i/" (Delattre et al., 205).

In other words, the two formants of the front vowels stand so clearly apart that they cannot be averaged. In the case of /i/, the one-formant version is *not* a result of averaging but is due to the fact that the high second formant can cue no other vowel. The ear may seize on it without going into further niceties. There is, then, a continuum of decreasing effort required for processing the various vowels. This continuum can be parsed into three clear stages: first, in back vowels, perceiving a fused acoustic signal and reanalyzing its formant frequency into its constituent formants; second, in front vowels, perceiving and discriminating two well-articulated formants; third, with /i/, perceiving a single formant without further discrimination. This assumption is corroborated by the effect of reducing the intensity of the lower formant of the back vowels, which caused a change in color toward the adjacent vowel. "We should suppose, then, that reducing the intensity of the lower formant . . . would have the effect of increasing the higher formant's relative contribution to the 'mean' and would thus effectively raise the mean formant" (ibid., 203). When, however, the lower formant's intensity was reduced in the front vowels, "the vowel color was replaced by a nonvowel sound. . . . In no case did a reduction in the intensity of the lower formant cause a clear *shift* in vowel color" (ibid.).

It is worth noting that the perception of *both* front and back vowels is subject to what the Haskins investigators came to call "categorical perception" so characteristic of the speech mode. The difference is that the system more easily exits speech mode (and categorical perception) where front vowels are concerned. Where nonspeech sounds are concerned, no categorical perception and no averaging take place. With arbitrarily chosen two-formant patterns, which did not sound like vowels, "the two formants of the nonvowel pattern, however close together they may be, did not fuse into a single sound, but tended rather to be heard as two-component chords" (206).

We might conclude, then, that in the case of the fused formants of the back vowels, as well as in the averaged one-formant versions, the ear reanalyzes a lowly differentiated acoustic signal in terms of the frequencies that appear to have been fused in it. This may be the application, at a lower level, of a mechanism underlying the motor theory of speech perception mentioned earlier.

How does this bear on the analogy of the vowel continuum with the brightness continuum? The acoustic signal of the back vowels is of relatively low differentiation. The impression from two formants received by the ear is sufficiently indistinguishable from that of one formant to place it somewhere intermediate between them. I submit, then, that relatively low differentiation and relative darkness are similar phenomenally to a sufficient degree to warrant matching the back vowel extreme of the vowel continuum with the dark extreme of the brightness continuum. Some perceptual quality of the acoustic signal sometimes intrudes into the speech mode, creating the poetic mode.

A still further step is required to complete our thought-experiment. Relatively low differentiation in vowel perception is naturally matched with darkness when the two continua are perceived as analogous. When back vowels are frequent in verse lines that refer to dark colors, the dark potential of the vowels is realized by combination with the dark elements of the meaning. Readers may intuit that the sound is somehow "an echo of the sense," or that the sounds are somehow "expressive" of the sense. It would appear that the conventionally established metaphorical relationship between darkness and mystic obscurity, or such emotions as hatred, is mediating between vowel color and such meaning components. I wish to suggest, however, that the relatively undifferentiated perceptions associated with the back vowels may be *directly related to* the lowly differentiated perception associated with mystic obscurities, and there is no need for a mediating metaphorical concept. Likewise, in light of my analysis of emotions and emotional qualities elsewhere (Tsur, 1978), low differentiation is a characteristic of emotional (as opposed to rational) qualities. Combined with high energy level, it can be an important ingredient in such emotions and emotional qualities as anger or hatred. Thus, when back vowels are frequent in verse lines depicting hatred or struggle, they may be perceived as expressive, with no need, again, for a mediating metaphorical concept. The foregoing analysis of the structural relationship between back vowels, darkness, mystic obscurity, and hatred or struggle may be regarded as further reinforcing one of the basic tenets of this study: that "past experience" insufficiently explains certain metaphorical intuitions; that some metaphorical relationships originate in processes that are deeper than cultural conditioning; and that certain culturally conditioned metaphorical conventions are results rather than causes, and they came to reflect, through repeated cultural transmission, certain underlying cognitive processes.

I have introduced the article by Delattre and his collaborators (1952) to account for the ubiquitous intuition that back vowels are darker than front vowels. That introduction also could serve as a basis for explaining the positive correlation of back vowels with mystic obscurity and hatred or struggle in poetry. In the next section I shall attempt with its help to explain some additional correlations.

In my book on metrics I have applied to metrical theory the cognitive principle that in a sequence of two coordinate items, other things being equal, it is more natural for the longer one to come last; it involves a lesser burden on the perceiver's short-term memory (Tsur, 1977: 76). The same principle applies to idiomatic phrases, such as "high and mighty" rather then "mighty and high," "bed and breakfast" rather than "breakfast and bed," "part and parcel," "blood and thunder," etc. In an important sense, relatively dark vowels seem to be marked in a way that renders them equivalent to relatively long words. Thus, we have phrases of two coordinate words, the second of which contains a prominent syllable that is darker than the prominent syllable of the first word; for instance: "sweet and sour," "flesh and blood," "safe and sound," "high and low," "meek and mild," "tit for tat," even "Peter and Paul" (overriding relative length), and "back and forth" (overriding logical order). Likewise, we have a long list of varied reduplications, going from the lighter to the darker vowel rather than from darker to lighter: "seesaw," "Ping-Pong," "sing-song," "ding-dong," "chit-chat," "tip-top," etc. Phrases like "sick and tired," "bread and butter," "milk and water" use both modes of marking, and so does "milk and honey," even though the order of words has been determined in Hebrew. This is, of course, only a tendency, not a foolproof rule.

There seems, then, to be some sort of equivalence between word length and phonetic darkness, determined by their greater load on the cognitive apparatus. Our cognitive economy tends to relegate to the end of the phrase (or clause) anything that requires relatively great processing effort, whether owing to relative word length or the relative difficulty of discriminating between the formants that constitute the back vowel's acoustic signal. In this sense, back vowels occupy more mental processing space and are perceived as heavier, or slower, or, at any rate, relatively marked. This can offer one possible explanation for the relative frequency of dark vowels in verse lines referring to slow and heavy movements. I shall return to this issue at the end of chapter 2.

A Real Experiment

It has been suggested that we might conduct a "real" experiment [5] to find out whether subjects who perceive back vowels as darker than

[5] This part of the research was jointly conducted with Yehosheba Bentov, Ruth Lavy, and Hanna Lock.

front vowels do indeed also perceive them as more complicated. So, we have conducted an experiment in which subjects were requested to characterize /u/ and /i/ with four pairs of antonymous adjectives directly applicable to the precategorical sensory information.

Subjects. Subjects were 120 students at the Seminar Hakkibbutsim teachers' training college, who were asked by their teacher to stay after the lesson, to take an experimental test.

Stimuli. The test material consisted of a sheet with the vowels /i/ and /u/ (in two conditions) printed in Hebrew and a list of eight pairs of antonymous adjectives. The test was forced-choice: subjects were requested to characterize each vowel by the more suitable member of each pair. In one condition the vowels were printed in isolation, in the other as part of the nonsense syllables /pit/ and /put/ (the word *pitput* means "chatter" in Hebrew). The test was administered to sixty subjects in each condition. Initially, the two conditions were meant to constitute a pilot test to choose the one that would yield more significant results. As will be seen, the differences between the two conditions were insignificant; however, the *direction* and *consistency* of the differences were rather significant. Two groups of pairs of adjectives were used. Four pairs (DARK□BRIGHT, FAR□NEAR, BIG□SMALL, LOW□ HIGH) refer to perceived qualities regularly and quite consistently associated with these vowels. Four pairs (THICK□THIN, DIFFERENTIATED□ UNDIFFERENTIATED, SPACIOUS□DENSE, SIMPLE□COMPLICATED) were the test adjectives and were chosen because they may characterize the perceived quality of the acoustic information as discussed. This structure was chosen mainly to get around a problem of communication with the subjects. If you explain to them clearly what kind of intuition you are looking for, you may interfere with their spontaneous response; and if you refrain from such an explanation, you cannot be sufficiently sure that the responses reflect the *kind* of perceived qualities you are interested in. For this reason the first four pairs of adjectives were included. They could, on the one hand, specify to some extent the kind of task the subjects were expected to perform on the test items; on the other, they could indicate whether the subjects understood the task.

Four pairs of adjectives also were included in the second group, not for symmetry, but because of the uncertainty concerning which pair would describe the perceived quality of the acoustic information least ambiguously. The possibility of such an ambiguity has been confirmed, unexpectedly enough, in relation to one pair of adjectives in the first group (see below). In this respect, the present test was initially intended as a pilot to choose the appropriate pair of adjectives. The results, however, were surprisingly unambiguous (in the expected direction) with respect to all four test pairs.

A few comments on the adjective pairs used. The Hebrew adjectives for THICK□THIN do not have the ambiguity of the English adjectives. *Samikh* means "thick" as in the phrases "thick syrup" and "thick smoke"; *dalil* (meaning "sparse" or "diluted") is perceived as its proper antonym. *Meruwah* in the third pair is ambiguous, meaning "spacious" or "wide open"; so there was a danger that it might be referred to the articulatory gesture rather than to the acoustic correlates, but its antonym means clearly "dense, crowded, compact." There was a suspicion that the expected use of DIFFERENTIATED□UNDIFFERENTIATED might be unfamiliar to the subjects. Finally, the adjectives SIMPLE□COMPLICATED are too abstract, and so it might be doubtful whether the subjects applied them to the sensory qualities of the acoustic correlates of the vowels or to some other aspects. Thus, it was expected that the adjectives THICK□THIN should best reflect the subjects' intuitions concerning the vowel's sensory qualities. The results are given in table 1.

All pairs of adjectives except one were very significantly attributed more frequently in the expected way than the other way around. The only pair that showed a tendency opposite to expectation was HIGH□ LOW. This seems rather odd, since the intuition that the step from /u/ to /i/ is more upward than downward is fairly consistent. Two factors seem to have influenced these deviant results. First, the results drew attention to the fact that the relevant Hebrew adjectives are ambiguous, meaning either HIGH□LOW or TALL□SHORT. Second, the occurrence of this adjective pair after the pairs BIG□SMALL and FAR□NEAR may have influenced subjects to construe it as referring to relative size rather than relative pitch.

When we compare the results for the /i/v/u/ pair with the results for the /pit/□/put/ pair, we find exactly the same general tendency, as well as similar numbers for each particular pair; the dominant attribution is only insignificantly lower for /pit/□/put/, whereas the recessive attribution is correspondingly higher. The relation between the two curves can be seen in figure 5.

Such results can be explained in terms of Rakerd (1984) finding that "vowels in consonantal context are more linguistically perceived than are isolated vowels." In other words, the perception of vowels in consonantal context is more categorical, whereas in isolated vowels more precategorical information can be perceived; alternatively, the sensory information underlying a vowel, by virtue of which it is typically associated with certain perceptual qualities, varies from one consonantal context to another, owing to "parallel transmission," that is, owing to "the fact that a talker often co-articulates the neighboring segments of an utterance (that is, overlaps their respective productions) such that the acoustic signal is jointly influenced by those segments" (123).

We have been seeking experimental support for certain speculations in cognitive poetics, namely, (1) in many cases the perceptual qualities associated with certain speech sounds can be accounted for by assuming that in the poetic mode *some* of the rich precategorical sensory information underlying the speech sound does reach consciousness; and (2) the perception of rounded back vowels as darker than the unrounded front vowels can be accounted for by the fact that F_1 and F_2 of back vowels are less differentiated, closer together, than those of front

Table 1. The majority of subjects characterizes /i/ as "lighter, nearer, smaller, thinner, more differentiated, more spacious and simpler" than /u/.

phoneme	dark	light	n.r.	far	near	n.r.	big	small	n.r.	low	high	n.r.
/i/	6	53	1	14	45	1	6	51	3	33	25	2
/pit/	14	46	--	20	40	--	17	43	--	36	23	1
/u/	53	6	1	46	14	--	52	7	1	26	31	3
/put/	46	14	--	40	20	--	43	17	--	23	36	1

phoneme	thick	thin	n.r.	undif.	dif.	n.r.	dense	spac.	n.r.	simp.	comp.	n.r.
/i/	9	49	2	21	37	2	18	39	3	51	8	1
/pit/	13	47	--	26	33	1	25	35	--	50	9	1
/u/	50	9	1	37	22	1	39	19	2	8	51	1
/put/	47	13	--	33	26	1	35	25	--	9	50	1

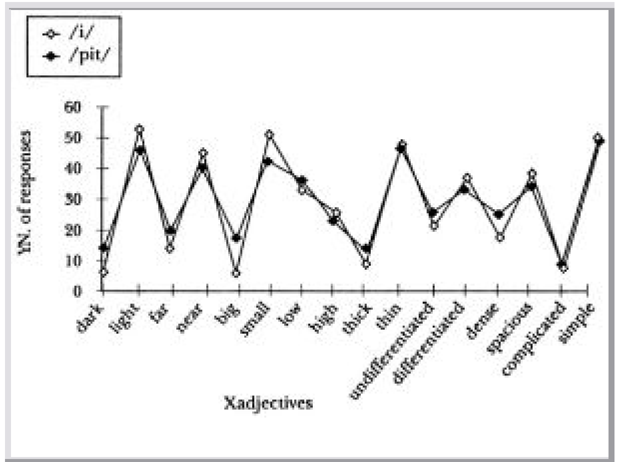

Figure 5.

Notice that in the pit curve the peaks and ebbs are consistently nearer to one another than in the i curve.

vowels. Our findings have been thoroughly consistent with these hypotheses. It is, however, questionable whether they constitute sufficient proof for them. Consequently, the validity of our "proof " crucially depends on the extent to which our findings fit with related facts and findings.

Angry and Tender Consonants

With the foregoing explanation in mind, let us return to Fónagy's data concerning the distribution of consonants in tender and angry poems. As was noted, the majority of consonants occur about equally in both groups of poems. I was wondering whether in some cases at least this equal distribution could not be attributed to the double-edgedness of certain sounds, that is, to the conflicting aspects of their combinational potential that cancel out each other's effects on statistics. There seems to be, indeed, some faint evidence to such mutual cancellation. The mean value of the frequency of /g/ in the tender poems of Hugo and Verlaine is 1.30 percent; in their aggressive poems it was 1.21 percent, an insignificant difference, indeed. However, the individual numbers for the two poets seem to be rather significant. /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 percent in his tender poems, and 1.35 percent in his angry ones. Likewise, the frequency of /d/ in the tender poems of the two poets is 7.51 percent, whereas in their angry poems it is 7.94 percent. But taken individually, again, the same sound has opposite emotional tendencies for the two poets. For Verlaine, its quality is basically tender (10.11:7.93), whereas for Hugo it has a basically aggressive quality (7.09:5.76). For Petőfi, both consonants are insignificantly more tender than angry. There are two possible explanations for these data. According to the first, these sounds have no emotional potentials of their own: the various poets may associate them in an arbitrary manner with any emotional quality, according to their idiosyncrasies. As these sounds recur, an emotional quality accrues to them according to their emotional contexts in each poet's work. The second explanation assumes that from the various acoustic, phonetic, and phonological features certain elements of the combinational potential may be abstracted, some of which may be conflicting.

It is difficult to give direct reasons for preferring any of these explanations, except that the second is more consistent with my analysis. This in itself is a not too strong reason that can be reinforced by a rewriting exercise performed by Hrushovski (1980:44):

Now, if this is the case, would not any sound pattern do? Let us try to "rewrite" the Shakespearean lines using words similar in content:

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

We have already created a very similar network of sounds, this time based on the repetition of *K*, strengthened by the cluster *K* + *R* (involving the original word "remembrance" too). Nevertheless, it seems that this sound pattern cannot possibly express silence, though "quiet thought" starts with *K* as "silent thought" starts with *S*. It is plausible that a reader will impute to this text something strong and harsh, reinforced by the sound pattern. The pivotal word may become "crux," though its counterpart "sessions" was subordinated to "sweet silent thought." One may generalize that, in a part of a text in which a sound pattern coexists with a number of semantic elements, the sound pattern may contribute to shifting the center of gravity from one direction of meaning to another.

Had the speech sounds no expressive potential of their own, the network of sounds based on **/k/** would have readily assumed the emotional quality of quietness, which it does not. Here the sound tends to confer on the text "something strong and harsh" and "may contribute to shifting the center of gravity from one direction of meaning to another" (for example, from quiet to crux). However, this putative shifting of the center of gravity became possible only through a regularization of meter in the transcription: the two successive, alliterating stressed syllables in "sweet silent" foreground these words and focus attention on their meanings. So let us amend the transcription to

When to the quorum of kind, quiet thought

also reinstating some of the legal connotations of the original text. But notice this: even though we are now prevented from shifting the center of gravity to some other "pivotal word," the /k/ sound retains its hard and strong quality and by no means becomes expressive of some "kind, quiet" atmosphere originating in the meaning of the words. The sound pattern becomes either neutral or improper to the emotion expressed. As this exercise may testify, speech sounds do have emotional potential of their own, and one may not ascribe to them just any quality suggested by the text's meaning.

Now, why is /k/ so hard and its repeated occurrence incompatible with an atmosphere of silence? And why are the voiced stops /d, g/ double-edged? If we relate these voiced stops to their voiceless counterparts /t, k/ that were positively correlated with aggression, and to the sonorants /l, m/ that were negatively correlated, we find that /d, g/ resemble the first group in one respect and the second group in another. As a preliminary, let us quote Fry's concise account (1970: 35 - 36):

The ear and the brain are quick to seize upon the difference between periodic and aperiodic sounds, between tones and noises, and can detect within very close limits the moment at which periodicity begins. In normal speech, all vowel sound, semi-vowels, liquids and nasals are periodic sounds, while noiseless consonants are aperiodic. Between these two classes, there are the voiced fricatives in which the ear recognizes an underlying periodicity, even though it is accompanied by aperiodic friction noise. In distinguishing between voiced and voiceless plosives, the exact moment at which periodicity begins is among the cues used by the listener.

Thus, the sequence vowels, liquids, and nasals, voiced fricatives, voiced stops, voiceless fricatives, voiceless stops, constitute a scale of decreasing periodicity or sonority, in this order. The feature [±PERIODIC] is responsible for the opposition TONES□NOISES, which is analogous, in a sense, to the opposition HARMONIOUS□NOT HARMONIOUS, or SOFT□HARD. There are two more relevant oppositions in the scale: CONTINUOUS□ABRUPT, and ENCODED□UNENCODED. The latter is a scale of relative encodedness rather than a dichotomy, and one should bear in mind that even the least encoded speech sounds are very much encoded. The optimal tender sounds are periodic (voiced), continuous, and relatively unencoded; the optimal aggressive sounds are aperiodic (voiceless), abrupt, and highly encoded. Now it should be emphasized that two points (or even two areas) on this scale can be picked out and presented in opposition to each other as more or less aggressive; liquids and nasals are not inherently tender, and voiceless stops are not inherently aggressive. When we see in Fónagy's table (197) that a sound like /d/ or /g/ is aggressive for one poet and tender for another, even in the same language, this may suggest that Verlaine treated /d/ as an abrupt and encoded sound and opposed it to continuous and relatively unencoded sounds, subduing its voiced feature; whereas in /g/ he subdued the abrupt and encoded features and opposed it to /k/ along the [±VOICED] feature. Hugo did just the reverse. We must grant the poets a considerable degree of free choice within the constraints of the sounds' combinational potential. Hence, the double-edged nature of these sounds.

The order of items on the scale is not unambiguous, precisely because the features vary independently. I have ordered the items according to the periodicity feature. According to the continuity and encodedness features, voiceless fricatives ought to have preceded voiced stops. Since, however, what matters is that owing to these conflicting features the sounds are double-edged, and can be separately contrasted to the other sounds on each of these features, their exact order in the middle of the scale seems insignificant. It should be noted that /r/, although a periodic liquid, has outstanding aggressive potentials, especially in languages in which it is rolled or intermittent. It is actually double-edged: periodic, but multiply interrupted.

One would have expected /b/ and /p/ to behave in a way similar to the voiced and voiceless stops, respectively; but no such evidence exists in Fónagy's tables. Possibly, these two bilabial stops have additional (nonemotive) expressive potential that obscures the statistical effects of the expressive potential shared with the other stops. There seems to be such a possibility. In his discussion of the child's gradual acquisition of the arbitrary linguistic sign system, beginning with the first syllable /pa/, and ending with the whole phonological system of his mother tongue (see chapter 2), Jakobson (1968: 72) observes that "nasalization is especially charged with emotion in the child." "The oral stop, on the other hand, carries either less emotion or no emotion at all, and is not used for complaining, but for 'drawing attention, dismissing, refusing,' and as a calmer, more apathetic designation, and thereby signals the real transition from emotional expression to symbolic language" (73).

Since /pa/ is the first syllable acquired by the child for use as an arbitrary referential sign, and it contains the bilabial /p/, it is not implausible that precisely the bilabial stops are more conspicuously associated with these nonemotional moods, in addition to the other aspects of the expressive potential shared with the other oral stops. As suggested in chapter 2, the later the acquisition of a speech sound, the greater its potential for being used in onomatopoeia or emotive sound patterns.

It seems, then, highly plausible that the ENCODED□UNENCODED Opposition is analogous in some way to the HARD□SOFT or RIGID□FLEXIBLE opposition, and that encoded corresponds to "hard" or "rigid." We should, however, again ask the question, why not match encoded with "soft," "flexible"? What perceptual qualities go with the ENCODED□ UNENCODED opposition? Liberman et al. (1972) describe a series of experiments by Crowder and Morton (1969), who found that in auditory (but not visual) presentation, vowels produce a recency effect in certain cognitive tasks, but stops do not. Part of the explanation seems to be as follows:

The special process that decodes the stops strips away all auditory information and presents to immediate perception a categorical linguistic event the listener can be aware of only as /b, d, g, p, t, or k/. Thus, there is for these segments no auditory, precategorical form that is available to consciousness for a time long enough to produce a recency effect. The relatively unencoded vowels, on the other hand, are capable of being perceived in a different way. . . . The listener can make relatively fine discriminations within phonetic classes because the auditory characteristics of the signal can be preserved for a while. . . . In the experiment by Crowder, we may suppose that these same auditory characteristics of the vowel, held for several seconds in an echoic sensory register, provide the subject with rich, precategorical information that enables him to recall the most recently presented items with relative ease.

This passage draws attention to several aspects of the ENCODED□ UNENCODED opposition that can explain its perceptual corollaries. First, relatively unencoded sounds appear to have some kind of sensory richness that highly encoded sounds lack. This intuitive observation can be explained by the assumption of Liberman and his colleagues that in the relatively unencoded sounds, the auditory characteristics of the signal, "the rich, precategorical information," can be preserved for a while in a "sensory register," whereas "the special process that decodes the stops strips away all auditory information and presents to immediate perception a categorical linguistic event." Several observations follow from this analysis. First, Hrushovski (1980: 39) mentions the insight formulated by I. A. Richards that poems are written with the "full body" of words; relatively unencoded sounds enable the language user to perceive a "fuller body" of words. Second, this sensory richness of the relatively unencoded sounds enables the listener to make fine discriminations within phonetic classes. Third, all this increases the ease with which certain cognitive tasks are performed (not only the highly artificial laboratory tasks that produced the recency effect in Crowder's experiment, but, possibly, also the task of reading; see Liberman and Mann, 1981: 128-29; Brady et al., 1983: 349-55; Mann, 1984: 1-10).

Now, the perception of rich sensory material and the making of fine distinctions precisely presuppose that openness, responsiveness, and susceptibility of adaptation characteristic of tender feelings and which seem to be lacking in aggressive behavior. Remaining firm in one's reliance on clear-cut categories and a lack of perceptiveness are important ingredients in a rigid and aggressive behavior. Reliance on rich precategorical information makes greater emotional and behavioral adaptability possible. It should be noted that Gestaltists (for example, Ehrenzweig, 1965) have observed that in visual designs clear-cut shapes assume a high degree of "plasticity" when perceived against a background of shadings, "inarticulate scribbling," or, in the present terminology, rich precategorical information. This is particularly true when gestalt-free sensory material is subliminally perceived. The richer this precategorical sensory information, the greater the plasticity of the consciously perceived visual shapes. Once the inarticulate scribbling is rich enough to reach consciousness, the peculiar conscious quality of plasticity disappears. The same seems to be true, mutatis mutandis, of the sounds of the great masters of the violin. This underlying mechanism can explain the tender plasticity perceived in relatively lowly encoded consonants, which are, in fact, clear-cut linguistic categories perceived against a background of rich precategorical sensory information.