Chapter 3

The Empirical Rationale

A rhythmical performance accommodates the conflicting patterns of stress and metre, and is half way between readings that suppress the linguistic stress pattern and those that suppress the metric pattern. A further distinction must be made *within* the scope of the rhythmical performance between "convergent" and "divergent" delivery styles, where both patterns are accommodated, but have different relative weights. This chapter is devoted to the nature of the rhythmical performance of poetry, and to the difficulties concerning its empirical study. It will admit that such an empirical investigation is *impossible in principle,* and will consider ways of circumventing the problem. It will be pointed out that the vocal devices uncovered in course of this inquiry are much more essential, much less dispensable in the divergent than in the convergent delivery style. Contrary to exposition conventions, I shall indicate the "empirical background" only after I state my own position. This will allow a more parsimonious comparison.

The Empirical Dilemma

This study assumes that the performance of poetry is not just a mediating agency between text and consumer; it is an essential dimension of any experiencing of poetic rhythm. It is conceived of here as of a solution to a perceptual problem posed by the versification patterns of poems and the stress and intonation patterns of their language. Therefore, the rhythmic performance of poetry should be accorded much greater attention than it usually receives and, I should add, of a different kind.

As I indicated in Chapter 1, all criteria for metricalness hitherto offered were violated by the greatest masters of musicality in English poetry, Milton and Shelley. The perception-oriented theory of metre performs a small Copernican revolution, and instead in the verse structure, it places the constraints in the reader's "rhythmic competence": the utmost limit of rhythmicality is the reader's ability or willingness to perform the verse line rhythmically. Jay Keyser of MIT (against whose generative metrical theory I was arguing, among others) suggested to me that such a formulation requires a systematic theory of "rhythmical performance". Wellek and Warren argue in their *Theory of Literature* (1956, Chapter 13) that in order to account for poetic rhythm, one must assume the existence of not one, but three metrical dimensions: prose rhythm, metric pattern, and performance (generative metrists have reinvented the first two of them). It would appear that Wellek and Warren need the performance dimension in order to account for the fact that two unlike delivery instances may still be performances of the same metric structure, and to point out that

Click here to go to the sound files for this chapter:
http://www.tau.ac.il/~tsurxx/Rhythm_Book_mp/Chapter3_Sound_Files.html

some "sound-recorders" mistake in their analysis an accidental performance for the poem's metre. For my purpose, performance is a perceptual solution to a perceptual problem, and as such determined by it to a considerable extent, but also leaving room for considerable creativity on the performer's part.

Let us begin with two issues extracted from a recent "state-of-the-art" summary of performance, in the "Performance" entry of *The New Princeton Encyclopedia of Poetry and Poetics* (1993). The first issue concerns delivery style: "C. S. Lewis once identified two types of performers of metrical verse: 'Minstrels' (who recite in a wooden singsong voice, letting scansion override verse) and 'Actors' (who give a flamboyantly expressive recitation, ignoring meter altogether)" (893). I shall argue that in between these two delivery styles there is a third one, which I call "rhythmical performance"; one, which is at the very core of poetic rhythm. The other issue concerns ambiguity. "Chatman isolates a central difference between the reading and scansion of poems on the one hand and their performance on the other: in the former two activities, ambiguities of interpretation can be preserved and do not have to be settled one way or the other ('disambiguated'). But in performance, all ambiguities have to be resolved before or during delivery. Since the nature of performance is linear and temporal, sentences can only be read aloud once and must be given a specific intonational pattern. Hence in performance, the performer is forced to choose between alternative intonational patterns and their associated meanings" (ibid. cf. e.g. Chatman: 1965, 1966). I shall argue that this is not so. I shall also argue that the two issues are intimately related. In Wellek and Warren's terms, the Minstrel subdues prose rhythm, and foregrounds the metric pattern; the Actor subdues the metric pattern in favour of the prose rhythm. For Chatman this may be a slight exaggeration, but in principle this is how things are and should be: when prose rhythm and metre conflict, "the performer is forced to choose between alternative intonational patterns". My position is that there is a third, "rhythmical performance", in which both metric pattern and linguistic stress pattern can be accommodated, such that both are established in the listener's perception. The same holds true of the conflicting intonation patterns articulating the linguistic unit (the phrase or sentence), and the metric unit (the line). This is precisely what the perceived rhythm of poetry is about, and by no means a side issue.

In spite of the enormous prestige of Wellek and Warren's book in general, no theory of performance has been attempted before my book. Fowler (1966), for instance, quotes their threefold distinction; but—very insightful, though—as its title "Prose Rhythm and Metre" suggests, his essay is devoted to only the first two dimensions. So I had to begin from scratch. I began with the early work of the so-called sound-recorders in poetry readings, but I mostly relied on introspection, listening to recorded readings by professional actors, attempting to explain my intuitions in terms of Gestalt psychology and speech research conducted at the Haskins Laboratories.

When I wrote in the early seventies to D.B. Fry of London University, seeking his help in my attempt to obtain empirical support for an instrumental analysis for my fledgling theory of rhythmical performance of poetry, he answered:

> It is clear to me that the ear and the brain are the only satisfactory instruments for dealing with matters of rhythm and metre. Not because their effects are any more "subjective" than any others, but because they represent a complex of impressions for which ear and brain are the only appropriate transducing instruments. So in fact I think you are better off listening to and thinking about metre than in doing physical measurements. [...] All this is not to say that one cannot learn something from physical measurements, but in the case of metre this is very much a matter of following one's hunches" (personal communication, 22 November 1972).

And that is what I did at the time: listening to recorded readings with my own ears, so that most of my early work was speculative. With the novice's enthusiasm I construed Fry's answer as if it meant that it was only a matter of time that electronic instruments would become available for this kind of research too. I was first initiated into the mysteries of speech research in 1980, when Al Liberman generously invited me to explore my theory on the equipment of the Haskins Laboratories in New Haven, and Terry Halwes guided me in the use of the Laboratories' facilities. But the state of the art still was not quite ready for my research, and I made very little progress. The problem through all these years was that there is an enormous discrepancy between what we experience when listening to speech and what can be read off a spectrogram. In fact, as Fry implied, this discrepancy is not due to the machines' incapability of representing the speech signal, but to the complex processing of the signal by the human brain. The sophisticated electronic instruments do give an accurate analysis of the sound information; but what really matters is its integration that takes place in the brain. Since then, every few years I applied to universities that had advanced phonetic laboratories for help in my research. Usually I received the answer that they had no staff experienced in this kind of research. Eventually I ended up at the Dpt. of Linguistics and Modern English, Lancaster University; there I was exposed to British phonetician Gerry Knowles's work, who quite independently from my plight created precisely the tools which I needed. The "state of the art" has not greatly changed during this period; it is rather that I found a way to ask the machine questions the answers to which make sense *within the framework of my theory.*

The view propounded here is essentially mentalistic, that is, it embraces the doctrine that mental states and processes exist independently of their manifestations in behaviour and can explain behaviour. This posed the greatest problem to my proposed empirical research. We have no access to what happens in that black box, the reader's head; we have access only to some external behaviour, that is, some vocal performances. We can only make inferences from these performances to mental pro-

cesses. And the vocal performances reflect the constraints of three kinds of competences, each later one relying on the preceding one: the competence to identify the conflicts between stress pattern and metre; the competence to find a solution to the conflict, and the proper command of voice to carry out the solution. When the performance of a deviant verse line is judged rhythmical, we may assume that the reciter had command of all three competences; when not, we may make only more or less accurate guesses as for which one(s) of the competences failed. This involved me in the need to find appropriate "informants", and then make judgments about every delivery instance whether it did or did not solve the problem.

It was obvious from the beginning that perhaps the most substantial part of experiencing poetic rhythm is outside the reach of the instruments. It has to do with a vocalizable linguistic string and its correspondence or lack of correspondence with an abstract mental pattern. From this relationship tensions and resolutions of tensions may arise. But from all this, only the linguistic string, if vocalized, is accessible to empirical measurements. When I got the big chance of my life, and the facilities of the Haskins Laboratories were at my disposal, the most important single piece of knowledge I received was that though you can make with those instruments quite sophisticated analyses of speech, not of those aspects which are of immediate relevance to metrical analysis. One can make very minute measurements of duration; but, as I shall argue at length in Chapter 9, measurable time has very little to offer directly to illuminate poetic rhythm: duration is an acoustic cue for stress or discontinuation, but poetic rhythm is not based on equal or proportional time periods. It is based, rather, on an abstract pattern that somehow exists in the perceiving consciousness; and perceived stresses sometimes confirm it, and sometimes disconfirm it. But, when I tried to find out what the machine can teach us about stresses, I was told that there is no way to infer from the machine's output what is the relative perceived stress of, e.g., two consecutive syllables. Perceived stress arises (as Fry's 1958 experiments prove) from an interplay of four kinds of acoustic cues (and their interaction with the listener's grammatical knowledge): intonational inflection; pitch change, duration and amplitude—in this descending order of relative effectiveness. There is no way to predict what the relative weight of each one of these elements will be in the resulting perceived stress. In my theoretical work I was also speculating about the effect of musical key and intervals on the perceptual articulation of the abstract mental pattern. I assume (see below, Chapter 9) that a cadential relationship between pitches can help to articulate certain metrical boundaries. So I hoped to get information about this from the machine. But I was told that there is no way to infer from the machine's output the perceived pitches on which words are uttered, let alone the perceived musical interval between them. In such questions, one must rely on the judgments of musically trained persons. The computer can be used only to prepare stimuli for such experiments (to isolate or reintegrate words or syllables in the original stretch of speech; or, for that matter, to excise some stretch of speech from one utterance and paste it into another one and then present it to observers' judgment, as will be seen below in Fodor *et al.*'s experiments). Over-articulation

too is an essential part of my theory of performance. But I was told that spectrograms can show the phonemes uttered; but not whether they are over-articulated.

As long as I was trying to obtain evidence for perceived stress and perceived pitch intervals, my empirical research was hopelessly stuck. Once I attempted to look, rather, for cues for tone-group boundaries and pitch movements, suddenly a solution appeared to be feasible. I had to restructure my quest and to formulate my questions in terms that the machine understands and in terms of which it can give significant answers. As I found out later, some of these answers may look quite trivial in isolation. They become significant only within an appropriate theoretical framework. Fortunately, my 1977 theory appeared to be tailor-made for just that. The big breakthrough occurred when I reformulated my questions in terms of continuities and discontinuities. This happened on my encounter with Gerry Knowles at the University of Lancaster. Two short papers by him contained all the information I needed for the breakthrough. When I came there, some very important work had already been done by Tom Barney. In his masters' thesis, an empirical study of the enjambment, he found ample empirical support for my speculations in the early ’seventies (contrary to the received view formulated by Chatman) that when the endings of the syntactic unit and the metric unit do not coincide (that is, when syntax is run on from one line to the other), the reciter may indicate continuity and discontinuity at one and the same time by having recourse to conflicting cues. This he did without having heard of my work before. I too started at this point. But then I discovered that other issues too, such as the rhythmical performance of consecutive stresses and of stress maxima in weak positions could be handled in those terms—within the framework of my perception-oriented theory of metre. In this chapter I shall give an example of each one of these.

The Empirical Solution

In one of his papers, Knowles (1991) investigated the nature of tone-groups. He distinguished internally defined prosodic patterns and external discontinuities at the tone-group boundaries. The former consist in some consistent f_0 pattern used in ordinary speech; the latter are temporal discontinuation (pause), pitch discontinuation (a sudden change in F_0) and segmental discontinuation (that is, in normal speech the articulation of adjacent words is overlapping; when there is no overlap, it may count as discontinuity, even if there is no pause). Glottal stops in words beginning with a vowel, or word-final stop releases too may indicate segmental discontinuation. This would be the most evasive type of discontinuity. “The important distinction that seems to be emerging is between boundaries with or without pauses”. In what follows, I shall explore how these correlates of tone-group boundaries can be exploited as conflicting cues for the perceptual accommodation of the conflicting patterns of speech and versification.

One of the most conspicuous kinds of segmental discontinuity is the prolongation of a phoneme or of a syllable at the end of an utterance, announcing (very much like *fermata* in music) that the preceding unit has come to an end. While this is most useful in the kind of research I am engaged in, there is a big problem with this notion. There is no standard by which we can determine whether a phoneme or sequence of phonemes is longer or shorter than ought to be. Consequently, one must rely in this respect on one's intuitive judgment, or some roundabout reasoning about measurements and comparisons. Prolongation is, in fact, a double-edged phenomenon, that is, in different contexts it has different, sometimes even opposite, effects. From a perceptual point of view, prolongation indicates lack of forward movement. Therefore, when we have reasons to suppose that it occurs at the end of some perceptual unit, it will be perceived as reinforcing the sense of *rest;* when it occurs in the middle of some forward movement, it is perceived as an *arrest*, arousing strong desire for change.

One more issue must be faced. Can we credit the non-phonetician listener with utilizing such evasive phonetic cues as phoneme prolongation, glottal stop, or the absence of coarticulation in the processing of the rhythmical performance of poetry? There is compelling evidence that listeners are heavily relying on their linguistic knowledge in perceiving either linguistic stress, or continuities and discontinuities in the stream of speech ignoring, sometimes even overriding, phonetic cues. This may cast some doubt on the efficiency of "allophonic" manipulations of phonetic cues in the rhythmical performance of poetry. Using French and Hungarian judges, phoneticians and non-phoneticians, listening to a Hungarian prose passage, Kassai and Fagyal (1996) found evidence to this effect; but also to the effect that tampering with the stream of speech (by the omission of silent periods) may increase reliance on the more evasive phonetic cues, as pitch discontinuity or such segmental discontinuities as phoneme prolongation, glottal stop, or the absence of coarticulation. This may suggest that over-articulation and the reduction of speech rate in poetry recital serve, among other things, to force these phonetic cues to the listener's attention.

In another paper, Knowles (1992) explores the alignment of the f_0 contour with vowels and consonants.

> Although the effect of a tone might be to highlight a whole word or phrase, its focus is on a single syllable. Within the syllable it focuses on the vowel, and if the vowel is a diphthong, on one of the elements of the diphthong. Ultimately within the relevant vowel there is a single point which appears to be the focus of accentuation (Knowles, 1992: 294).

Knowles calls this point the *accent point.* Such points may be located in various places in the vowel. Accordingly, he speaks of early-peaking and late-peaking, as the case may be. "Peak position would seem to be a continuous variable" (ibid). For our present interest it is important that peak position may affect the grouping of sylla-

bles. The results of his research are completely independent from the needs of the present inquiry. Knowles suggests the possibility that behind the phonological contrast of tone there is a functional contrast between an "initial" marker and a "final" marker.

> Looking at all final tones before different kinds of boundary it was found that the greater the boundary the earlier the peak, with the earliest peaks preceding the silent pauses forming a major tone group boundary. The position with regard to initial position is not so clear: the latest peaks follow [...] pauses [...] *within* a major tone group [...]. Tones at the beginning of a major tone group are neither early nor late (Knowles, 1992, 296; my italics — R.T.).

Perceptual Forces

I shall attempt to account for the effects of early and late peaking by what the Gestalt Theorists call "perceptual forces". This notion is based on the Gestalt assumption that a perceptual unit tends to preserve its integrity by resisting interruptions. I shall introduce the issue by way of Arnheim's explorations of the phenomenon in visual perception, and then discuss it regarding speech perception, on three levels: on the subphonemic level (i.e., early and late peaking), on the syntactic level, and on the versification-units level. By this I wish to suggest that certain phenomena at various levels of poetry may be handled by a homogeneous set of principles, inherent in the human brain. Only the first two levels are treated experimentally; the third one will be treated sepculatively only, in the terms "Articulateness" and "Requiredness".

Two aspects of Knowles's foregoing discussion of peaking is of interest for our present inquiry. First, late peaking at the beginning of a tone group tends to have a forward rather than a backward grouping effect. And that it tends to be associated with minor rather than with major tone group boundaries. Knowles's research on this issue is at its very beginning; so, these findings must be handled very carefully. But the finding concerning the tendency for late-peaking to begin rather than end a group may gain some support from Gestalt theory. We may be up against what Arnheim called *perceptual forces.* These forces are supposed to be inherent in both visual and aural perception. The existence of perceptual forces is most conspicuous in visual perception and this is the domain primarily explored by Gestalt psychologists. At the beginning of his 1967 book, Arnheim demonstrates "the hidden structure of a square" by placing a black cardboard disk in various positions on a white square. Thus he "maps out" regions of tension and of balance. In figure 1, the disk lies slightly off the centre. "In looking at the disk we may find that it does not merely occupy a certain place but exhibits restlessness. This restlessness may be experienced as a tendency of the disk to get away from where it is placed or, more

specifically, as a pull in a particular direction—for example, toward the centre" (Arnheim, 1967: 2).

Psychologically, the pulls in the disk exist in the experience of any person who looks at it. Since these pulls have a point of attack, a direction and an intensity, they meet the conditions established by physicists for physical forces (ibid., 6). Although perceptual forces are not physical in the sense that gravity is, "there is no point in calling these forces 'illusions'. They are no more illusory than colours, which are attributed to the objects themselves, although they are actually nothing but the reactions of the nervous system to light of particular wave lengths" (ibid., 8).

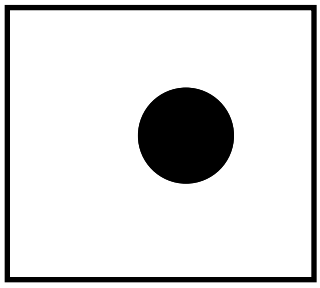

Figure 1

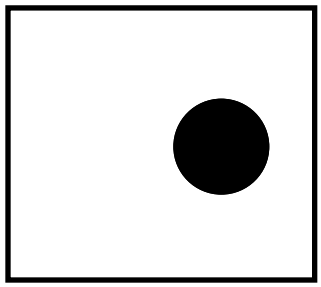

Figure 2

The skeptical reader may object: "Let us grant that these 'perceptual forces' do exist in visual patterns (or even in musical patterns). But verbal expression may be quite different. It introduces referential meaning through an arbitrary system of linguistic structures. How can we know that your 'perceptual forces' survive in such a complex system, too?" It is difficult to give a conclusive answer to this objection. It is a most commendable and wise precaution in any inquiry not to take anything for granted. However, when pursued to the extreme, this principle may be crippling. Mentalist approaches to art are particularly vulnerable to such skepticism, since we know very little about the mental processes involved. Since in literary criticism it is hardly possible to go beyond the accumulation of unrelated facts without appealing, to some extent, to the critic's intuitive perceptions or to a theoretical framework, or to both, I shall make a brief attempt to meet the above objection by referring to psychological experiments by way of testing the notion of constituent or phrase structure of sentences, performed by Fodor, Bever and Garret at MIT.

> These researchers devised an ingenious technique for revealing the presence of phrase boundaries in the perception of sentences. The technique is based on the Gestalt assumption that a perceptual unit tends "to preserve its integrity by resisting interruptions" (Fodor and Bever, 1965: 415). In the experiment of Fodor and Bever, subjects listened to a sentence during which a click occurred, and immediately afterward were required to write down the sentence and indicate where the click had occurred. If a phrase is a perceptual

unit, subjects should tend to hear a click which occurred during a phrase as having occurred between the phrases.

One of their sentences was "That he was happy was evident from the way he smiled". This sentence has a major break between "happy" and "was". A click was placed at various positions in this sentence. […] Each subject heard the sentence with only one click on it.

Fodor and Bever found that subjects were most accurate in locating the click which occurred between the two major phrases of the sentence—i.e., between "happy" and "was" in the above example. Clicks occurring before this break tended to be displaced towards the right (i.e., into the break), and those occurring after the break towards the left (i.e., again into the break). Fodor and Bever conclude that their findings "appear to demonstrate that the major syntactic break plays an important role in determining the subjective location of noises perceived during speech", thus supporting the hypothesis that "the unit of speech perception corresponds to the constituent". One might call these results into question on the suspicion that the major syntactic break is signalled by some acoustic means, such as pause. In additional research, however, Garret, Bever and Fodor (1966) have demonstrated that there are no clear acoustic cues that mark the breaks between constituent phrases. The most dramatic evidence of this surprising fact comes from an experiment comparing pairs of sentences such as:

(1) As a result of their invention's *influence the company was given an award.* * *

(2) The chairman whose methods still *influence the company was given an award.* * *

When subjects were asked where they hear the longest pause in the sentences, they report—as one might expect—that they hear a pause in (1) between "influence" and "the", and in (2) between "company" and "was". The perceived pause thus corresponds to the major constituent boundaries in the two sentences.

The ingenious part of the experiment comes next. The two sentences were recorded on tape, and the two italicized segments were interchanged. […] *Subjects' perception of pause location, however, was unchanged.* The same was true of click displacement. As indicated by asterisks in the two sentences above, a click occurred either during "company" or "was". The perception of click location, however, was significantly different for the two sentences. The click in sentence (1) tended to be heard between "influence" and "the", and in (2) between "company" and "was". But remember the sentences were acoustically identical (Slobin, 1971: 25-26; italics in original).

For our present purpose, there are two important implications of these results: first, that perceptual forces do exist in a linguistic environment; second, perceptual forces in a linguistic environment are crucially influenced by the intruding event's coincidence with, or distance from, the boundary of the perceptual unit. In prosody, however, there is a further complication. One cannot elicit perceptual forces with the help of some extra-linguistic click. The immediately observable exponents of all the perceptual units in poetry must be linguistic units. There are, however, two sets of organizational principles in poetry from which such perceptual units arise: one belongs to language, the other to versification. In one, we encounter a hierarchy of boundaries: sentence, clause, phrase, word and syllable; in the other, we encounter another hierarchy of boundaries: stanza, verse line, hemistich, and metric foot. When the major boundaries of the two sets coincide, there is a feeling of stability. When, however, they do not, the boundaries of perceptual units intrude upon each other, and perceptual forces of various intensities arise. Since, according to the gestalt assumption, a perceptual unit tends "to preserve its integrity by resisting interruptions", each one of the conflicting units will tend to reinforce itself in the reader's perception and, up to a certain point of complexity, their vividness will be heightened by the intrusion. It may well be that Knowles's finding suggests that these forces are active on the sub-phonemic level too. Indeed, the more examples we find in poetry reading, the more we become convinced of this.

Articulateness and Requiredness

I have already briefly mentioned these two terms. In the present section, I propose to explore some of their implications. *Articulateness* and *Requiredness* are two sides of the same coin; they are aspects of breaking up a whole into segments. When we speak of articulateness, we imply that a whole has been broken into parts, and that this facilitates perception of the whole. When we speak of Requiredness, we imply that each part is essential to the whole: when a part is omitted, there is an acute feeling of incompleteness, of imbalance. Articulateness and requiredness depend on the relative strength of the whole. Requiredness is possible only where the whole is highly organized. If the integrity of the whole is not felt, deficiency cannot be felt either.

I mentioned above that according to the Gestalt assumption, a perceptual unit tends "to preserve its integrity by resisting interruptions": each of two conflicting units will tend to reinforce itself in the reader's perception. The articulation of a perceptual unit suggests its interruption. Thus, articulation enhances requiredness in cases when the whole is strongly organized; at the same time, it also creates a need to reassert the integrity of this whole. The weaker the organization of the whole, the weaker the impact of requiredness. At the same time, the role played by articulation in perception increases and may become all-important. In a word, the smaller the

relative strength of the whole, the more the emphasis shifts from the requiredness aspect of segmentation to its articulateness aspect.

Articulateness in poetry involves a *double entendre*. The word refers to both the idea of "clear, distinct", and the idea of "jointed". Pope is a highly articulate poet (in the sense that his long poems are segmented into well-shaped couplets and these again into lines and their subordinate iambic feet):

1. Some foreign writers, some our own despise,
   The ancients only, or the moderns, prize.

The reader perceives here a distinct quality of wit, which seems to depend on the segmentation of the utterance. We do not perceive the whole as an undifferentiated lump of words, but break it up into small, easily perceptible units. There is a clearly defined couplet, which consists of clearly defined lines which are broken up into phrases. In the present example, the sentence converges with the couplet, the clauses reinforce the lines; in the first line, the break between the two phrases comes exactly at the middle. There is a similar break after the fifth syllable of the second line, but there is an additional, less expected break after the ninth syllable. We may observe that the nearer the interruption to the end, the greater its threat to the verse line's integrity as a perceptual unit, and the stronger the perceptual unit's resistance to interruption. At the same time, we can observe the requiredness of the last segment ("prize") in the second line, by trying to omit it. An exceptionally strong feeling of deficiency, of imbalance will follow.

When the syntactic unit (phrase, clause) does not converge with the line, it may, then, give rise to one of two opposite tendencies. The shape of the line may be blurred, or a tension between the syntactic and prosodic units may be generated. In the latter case, both syntactic and prosodic units "strive" to establish themselves in the reader's perception. So, in case of divergence, we may witness that the line is perceived as sharper than usual. Only the latter case will be briefly considered here.[1] If divergence goes beyond a certain point, the various shapes "lose control" and begin to fade into one another.

Let us consider a minimal pair, of a couplet from Pope's *The Rape of the Lock*, and an altered version with a different word order:

2. Now Lapdogs give themselves the rouzing Shake,
   And sleepless Lovers, just at Twelve, awake.

3. Now Lapdogs give themselves the rouzing Shake,
   And, just at Twelve, the sleepless Lovers wake.

The difference between the two versions seems rather slight. Still, quote 2 appears to be wittier. One of the salient reasons seems to be the difference in the place of the

[1] For the former, the reader is referred to Tsur 1972; 1992a: 139-148.

syntactic breaks. In 2, "And sleepless Lovers, just at Twelve" urgently *requires* completion. Psychologists of perception and personality speak of "leveling and sharpening" in a variety of mental processes. If figures that slightly deviate from symmetric patterns are presented under conditions that keep the stimulus control weak enough to leave the observer with a margin of freedom, some will perfect the symmetry of the model, whereas others will exaggerate the asymmetry so as to eliminate the uneasy feeling of ambiguity. The nearer to the end the break occurs, the stronger the "drive" felt to reach the end, and the greater our relief when the missing part is supplied. A break after the seventh, eighth, or ninth syllable generates increasingly greater tension, so that in quote 2, above, "awake" is highly required, and its occurrence is experienced as sudden relief—as wit.

Divergence of prosodic and syntactic units may exceed the single line. Enjambment is relatively infrequent in Pope's poetry. But when he resorts to it, it serves to enhance the sharpness of the couplet rather than blunt it. Consider the following couplet, for instance:

4. For Spirits, freed from mortal Laws, with ease
Assume what Sexes and what Shapes they please.

"With ease" is highly required to establish a strong line, with a sharp, closed shape. This line, however, achieves no closure. The clause runs on to the next line. Far from being blunted, the couplet "seeks" to establish and enhance itself. That is, distortion of symmetry and enjambement in the first line renders the second line of the couplet more required in order to yield a closed, stable shape. The closure of the couplet is reinforced by the parallelism and massive symmetry of the two phrases in the second line.

Perceptual units tend, then, to preserve their integrity by resisting interruption. This has two very important effects in poetry reading. First, perceptual units threatened by interruption tend to reinforce themselves in the reader's perception, up to a certain point. And second, the nearer the intruding event to a boundary, the stronger its "drive" toward that perceptual boundary. In poetry reading, no external mechanical clicks are supplied. On the subphonemic level, the intruding event may be an intonation peak, that may hit the phoneme at any point, in the middle, or somewhere between the middle and one of the phoneme boundaries. When the ends of the syntactic and the versification units do not coincide, then the end of the syntacic unit may "intrude" upon the verse line, or the end of the verse line may intrude upon the syntactic unit. In both cases, considerable perceptual forces will be generated. As will be seen in most chapters of this book, but mainly in Chapters 7 and 9, this perceptual dynamics may be all-important in poetry.

In my book (Tsur, 1977: 134 and passim) I suggested that where linguistic stress pattern or intonation pattern conflict with the patterns required by versification, the performer may have recourse to conflicting cues. This will apply, as I hope to show, both to enjambment and to instances where the linguistic stress pattern devi-

ates from the metric pattern. These instances may be explained in terms of the hierarchy of cues provided by Knowles, supplemented by his notions of early and late peaking. In the next three sections, I shall apply the method adopted here to three cases illustrating three issues: enjambment, sequence of stressed syllables, and stress maximum in a weak position. In all three issues the need to indicate continuation and discontinuation at one and the same time will be conspicuous; and I have chosen instances in which segmental discontinuation plays a substantial role.

First Issue: Enjambment

Enjambment is an obvious instance in which patterns of language and versification conflict. The sentence is run on from one verse line to the other. Consequently, the line ending requires the reader to stop, the run-on sentence requires him not to stop. To have his cake and eat it. This issue is focussed on a verse instance from Keats's "Ode on a Grecian Urn" in which the versification unit (the verse line) conflicts with the syntactic unit (the clause), that is, when the phrase or clause runs on from one line to the next one. In such a verse structure, versification requires that the reciter stops at the end of the line, whereas syntax requires not to stop; to have one's cake and eat it. I am going to compare two readings of the same verse instance by leading British actors, available on commercial records. We have listened to them and got the impression that one reading does solve this problem, the other one does not. I am going to examine the phonetic contrasts between the two readings which may account for this perceptual difference.

5. Sylvan historian, who canst thus express
A flowery tale more sweetly than our rhyme...

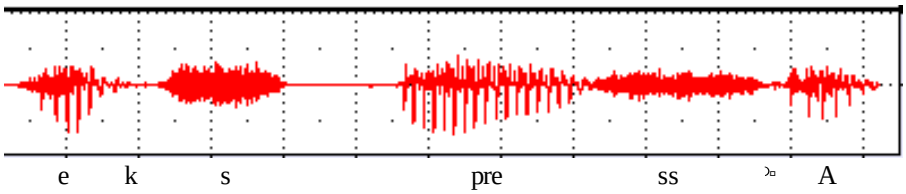

Figure 3 Wave plot of "express A" in Hodge's performance (ˀ indicates glottal stop)

Listen to sound file

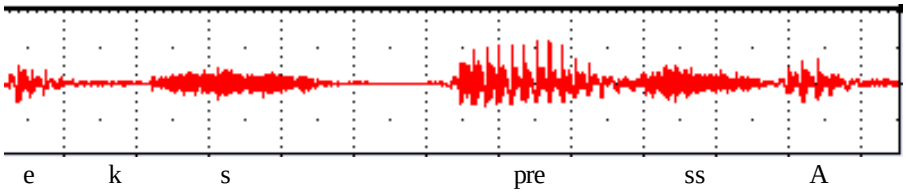

Figure 4 Wave plot of "express A" in Sheen's performance (no glottal stop)

Listen to sound file

Here the versification unit (the line) ends after the word "express", whereas the linguistic unit (the clause) does not end there, and runs on to the next line. It is assumed here that a rhythmical performance of these two lines will suggest continuity and discontinuity at one and the same time; a non-rhythmical performance will suppress either the continuous or the discontinuous aspect of this structure. How can we know whether a delivery instance displays at this point continuity or discontinuity or both? Paraphrasing Sibley (1962) we might say that we know that a delivery instance is continuous or discontinuous or both by listening, just as we see that the book is red by looking, or as we tell that the tea is sweet by tasting it. By listening to two delivery instances of this verse instance, we may prefer one to the other according to whether it does or does not suggest continuity and discontinuity at the same time. We can also establish the phonetic correlates that make these suggestions. The present approach assumes that continuity and discontinuity *can* be suggested at one and the same time by using conflicting phonetic cues, thus committing "organized violence" against speech processing. This cannot be done by merely looking at the graphic output of the computer, only by listening to the sound output. Then one may determine from the graphic output, which features of the speech signal "typically count toward" continuity, and which ones against it, to use again Sibley's terms. But only by listening we can tell what the perceived quality of the whole is, continuous or discontinuous or both.

An MA seminar group, my research assistant and myself have listened to commercially available recordings of Keats's Ode by two leading British actors, Douglas Hodge and Michael Sheen. We all made the judgment that Hodge offers an admirably rhythmical solution of the problem, by suggesting continuation and discontinuation at one and the same time at the end of the word "express", whereas in Sheen's reading "A" at the beginning of the next line is irritatingly continuous with "express". Then we were looking for features that typically count toward or against discontinuity. Unfortunately, owing to recording quality, the machine produced no pitch contour for "A" in either reading; so we were unable to use pitch movement as an indicator of continuity or discontinuity. At any rate, the overall pitch contour of the relevant segments seems to suggest continuation in both readings. There is no measurable pause in either of the readings between the two words; and this takes care of syntactic continuity. At the same time, there are two rather significant differences between the two readings that may account for the perceived difference between them. First, in Sheen's reading the /s/ of "express" is inseparably run into "A", whereas in Hodge's reading we may discern a glottal stop that perceptually separates the two words, indicated by a minute "lump" in the wave plot. (Glottal stop is the speech sound we insert before "aim" when we say: "I said 'an aim', not 'a name'". Second, a glance at the two figures may indicate that the syllable "press" in general, and the closing /s/ in particular are considerably longer in Hodge's reading than in Sheen's. There is no way to know whether a given phoneme in a stretch of speech is longer or shorter than ought to be; one may make only roundabout comparisons

of relative duration. In Hodge's reading /s/ is 150-msec-long, in Sheen's 105-msec long, that is, over 1.42 times longer. In Hodge's reading "pre-" is 1.15 times longer than in Sheen's (183: 159). In "ex-", by contrast, the /s/ is 111-msec-long in Hodge's reading, in Sheen's 112-msec-long. One msec difference is insignificant; but here this minute difference in Sheen's favour should be evaluated against the substantial, 1.42 times greater length of the final /s/ in the same word in Hodge's reading. The whole phrase "who canst thus express" is 1.450 sec long in Hodge's, 1.334 sec long in Sheen's reading, that is, only 1.08 times longer. That is, "-press" is longer in Hodge's than in Sheen's reading, *relative to* their respective contexts. One has a strong intuition that in Hodge's reading the line-ending is clearly articulated in spite of the run-on syntax, whereas in Sheen's reading it is not; and that this difference has to do with the relative duration of the /s/ (and of "press") in the two readings, and with the presence or absence of the articulatory gesture called glottal stop. The graphic output of the computer fully supports these intuitions. All available psychological and phonetic evidence supports this relationship between structure and intuition.

Second Issue: Consecutive Stresses

The application of conflicting cues indicating continuity and discontinuity is very obvious in enjambment. Their applicability to the solution of rhythmical problems arising from strings of equally stressed syllables is much less obvious. Here they may become significant only in the perspective of the cognitive assumptions expounded in Chapter 2. Mental processing space may be saved by the over-articulation of phonemes and of word boundaries. The linguistic sequence requires continuity; the over-articulation of word boundaries consists in discontinuity. The resulting extra processing space is required for the simultaneous perception of regularly alternating weak and strong positions that constitute the metrical set, and of the immediately observable string of equally stressed syllables. Consider the following line from Shakespeare's Sonnet 1:

6. But thóu, contrácted to thine ówn bríght éyes, ...
   w s w s w s w s w s

In the last three syllables we have a succession of three linguistic stresses, whereas the metrical positions they occupy constitute a sequence of strong—weak—strong. The stress of a pre-nominal adjective is subordinated to the stress of the noun it qualifies. Marlowe Society in their full recording of Shakespeare's Sonnets (Argo ZPR 254) perform the word *own* with a steep rise which immediately falls; then pitch resets to a higher level on *bright*, then it falls again to low on the syllable *eyes*. The result is that the pitch pattern drastically deviates from what might be expected in spoken English, yielding three consecutive, heavily stressed syllables.

I have predicted that the performance of such sequences as the last three syllables would display two opposite tendencies. On the one hand, they would be grouped more than usually together; on the other hand, the words themselves, as well as the boundaries between them would be more than usually well articulated. Knowles provides the prosodic distinctions that can be exploited for showing how the various acoustic and phonetic correlates suggest conflicting cues for the generation of the opposite tendencies. As Knowles pointed out, temporal discontinuation (pause) usually cues major tone-group boundaries.

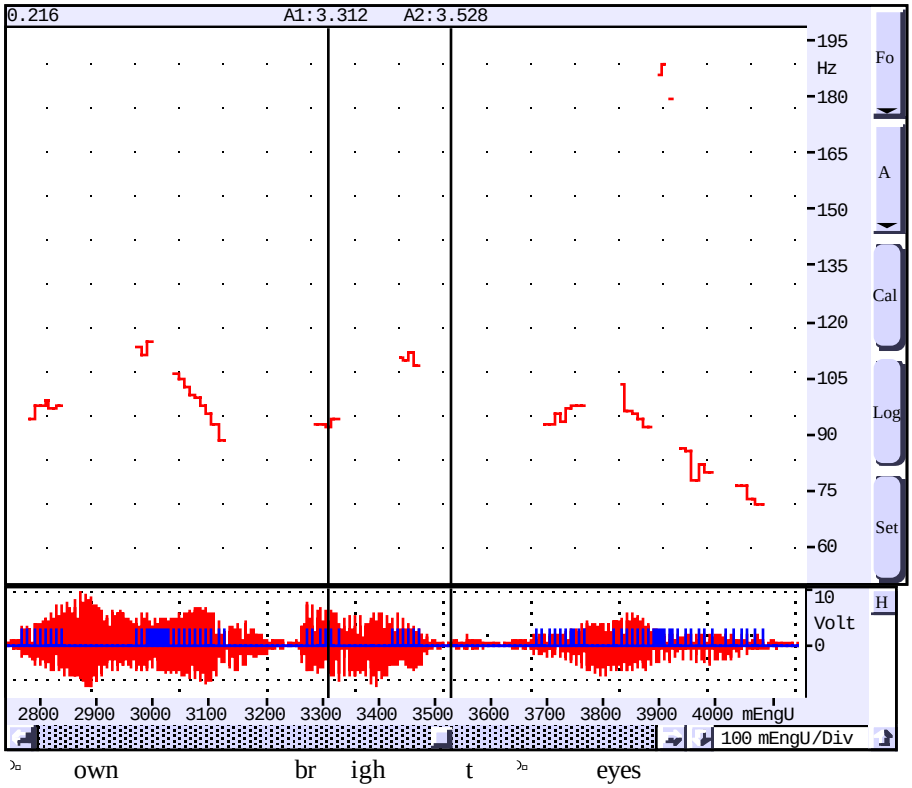

Figure 5 Wave plot and F_0 extract of "own bright eyes" in Marlowe Society's reading. The two markers indicate the boundaries of the diphthong in *bright*.

Listen to sound file

Marlowe Society perform the word sequence "contracted to thine own bright eyes" as some internally defined prosodic pattern: there is a continuously falling pitch pattern (with almost imperceptible but consistent differences) connecting the highest points of the stressed syllables. At the same time, they cue the over-articulated word-boundaries by cues other than pauses. One such conspicuous cue is the very unusual pitch-movements characterizing the utterance of each one of the three words. The other ones are very evasive, and are what have been called "segmental"

discontinuities. In regular speech the normal course would be that words like “thine own” or “bright eyes” are overlapped in articulation. However, a close inspection of the spectrogram revealed three very unusual features in such word sequences. At the end of *bright*, the /t/ sound is released; and *eyes* begins with a glottal stop. These features count as deliberate disruption of continuity. At the beginning of *own* there is another glottal stop which, too, is wholly unexpected in the flow of ordinary speech; and is preceded by a longer than usual /n/, which too indicates segregation. In this way, the last three words are segregated from the rest of the tone group; this, however, as well as the clear-cut articulation of word-boundaries does not interrupt the integrity of the tone-group predicted on purely linguistic grounds. The last three monosyllabics can be segregated from the rest of the syntactic unit, and grouped more than usually together; at the same time, their boundaries can be more than usually well articulated.

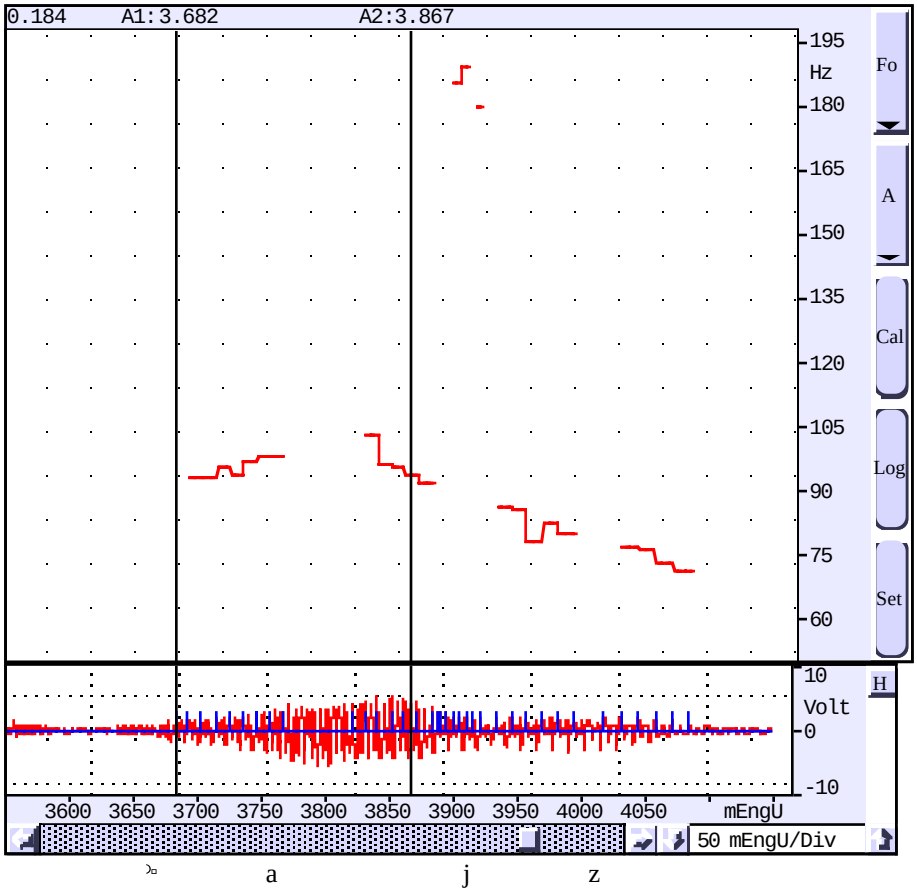

Figure 6 Wave plot and F_0 extract of “eyes” in Marlowe Society’s reading. The two markers indicate the boundaries of the /a/ of the diphthong.

When I wrote the first draft of this analysis, we were not yet aware that a major forward grouping agent in the rhythmic performance of divergent poetry is what Knowles called "late peaking". The peak of the pitch contour normally occurs in the middle of the syllabic crest; in some instances, however, it occurs late in the vowel, or even after it; and sometimes it occurs earlier than the middle. We have found in our corpus that late peaking generates an impetuous forward drive; in fact, the later the peaking, the more impetuous is the forward drive. On returning to the phrase "thine ówn bríght éyes", we have found that there is a late peak on each one of these stressed syllables, a moderately late peak on the stressed syllables in strong positions, namely "ówn" and "éyes", and an exceptionally late peak in the stressed syllable in weak position, namely "bríght".

Third Issue: Stress Maximum in a Weak Position

Regarding this issue, the relationship between the perceptual problem arising from the violating stress maximum in a weak position and the conflicting cues indicating continuity and discontinuity is even less direct. These conflicting cues become significant, again, only with reference to the information-processing model propounded in Chapter 2, that requires the saving of mental processing space for the simultaneous perception of the conflicting patterns of stress and metre. But there is here a greater number of intervening steps. The stress maximum in a weak position cannot be handled in isolation, but as part of a "stress valley" that violates metre in, e.g., the seventh (weak) position, and confirms it in the tenth (strong) position. According to Gestalt Theory, such a closed, symmetrical stress valley occupies relatively little mental processing space. Furthermore, as Gestalt Theory teaches us, greater simplicity may be achieved if this closed, symmetrical stress valley is segregated from the rest of the line, in the middle of a phrase. At the same time, the phrase, the syntactic unit, demands continuity across the segregating discontinuity.

Consider excerpt 7. It contains two deviances that render, according to three generative theories, a line unmetrical, by three different criteria. First, in the string of stressed syllables "bards gild" the last one bears the greater linguistic stress; but the string ends in a weak position. According to Bruce Hayes, such a construction renders a verse line unmetrical. Second, the syllable *lap-* is a stressed syllable between two unstressed ones, that is, a stress maximum, and occurs in a weak position. According to Halle and Keyser, such a construction renders a verse line unmetrical, even if it involves no polysyllabic. Third, according to Kiparsky's generative theory, when the stressed syllable of a polysyllabic occurs in a weak position, it renders the line unmetrical, even if it is not a stress maximum. Figures 7-8 show the reading of this line by DF, a colleague from the academy, who proved to be one of our most competent readers (in the terms defined above). In harmony with the predictions in Chapter 2, the consecutive stressed words are, in this reading (in all

readings, in fact), over articulated and their boundaries overstressed mainly by the falling, terminal intonation contours. The syllable *lap-* is perceived as over-articulated, over-stressed, segregated from the preceding article, and pushing forward, to reach stability at the last syllable.

7

7. How mány bárds gíld the lápses of tíme
w s w s w s w s w s

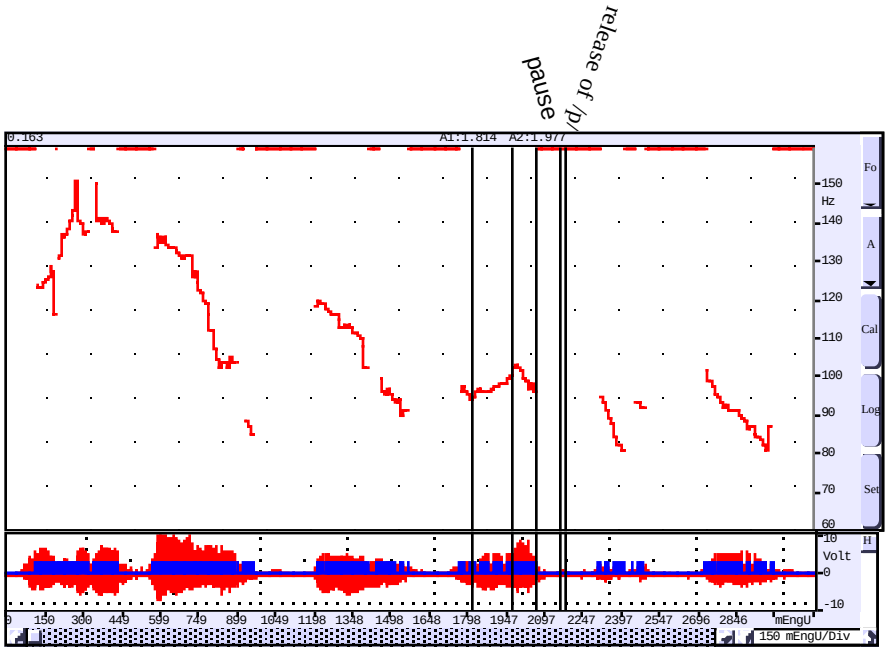

Figure 7 Wave plot and F_0 extract of "How many bards gild the lapses of time" in DF's reading.

Listen to sound file

In Chapters 1 and 2 I made predictions as for how a stress maximum in the seventh (weak) position would be performed. I suggested that it requires the foregrounding and segregation of such stress valleys as "lápses of tíme" in positions 7-10. When the structure of the whole line becomes too complex, too irregular, subdivision may yield better shape; in the present instance, the stress valley deviating from the iambic pattern is perceptually isolated, yielding a more or less iambic chunk plus an isolated closed, symmetrical stress valley. This involves the reciter in a problem of continuity and discontinuity. The stress valley begins in the middle of the phrase "the lapses", and requires segregation; but the phrase as a phrase requires continuation. There is no measurable pause between the two words. That takes care of continuity. Now the source of the segregation is rather evasive. As we shall see in

Chapter 6, usually some pitch discontinuity is involved in the segregation of such a stress valley. Here there is a rather mild rise of pitch on the /l/, followed by a small peak and change of direction. This could hardly account for the clearly audible effect of forward grouping on *lap-*, away from "the". One might improve upon this observation by pointing out on the wave plot that there is also a sudden rise of energy from the /l/ to the vowel. But the crucial factor appears to be a segmental discontinuation: excessive duration of the /l/. Consider the following durations of phonemes in "lapses": l: 162 msec; a: 89 msec; pause: 100 msec; release: 13 msec; s:120 msec; e :98 msec; s: 47 msec. There is an exceptionally long pause between the vowel and the release of /p/, perceived as over-articulation of the voiceless stop (cf. Chapter 9). And there is an excessively long /l/ at the word onset, indicating segmental discontinuity between the two words, pushing forward, to the last position. To test this perceptual judgment, I excised a stretch of 118 msec from the /l/, leaving a 44 msec /l/. The result was a perfectly natural /l/, the only difference being that *"lap"* became more stable: the sense of discontinuity and the forward drive disappeared.

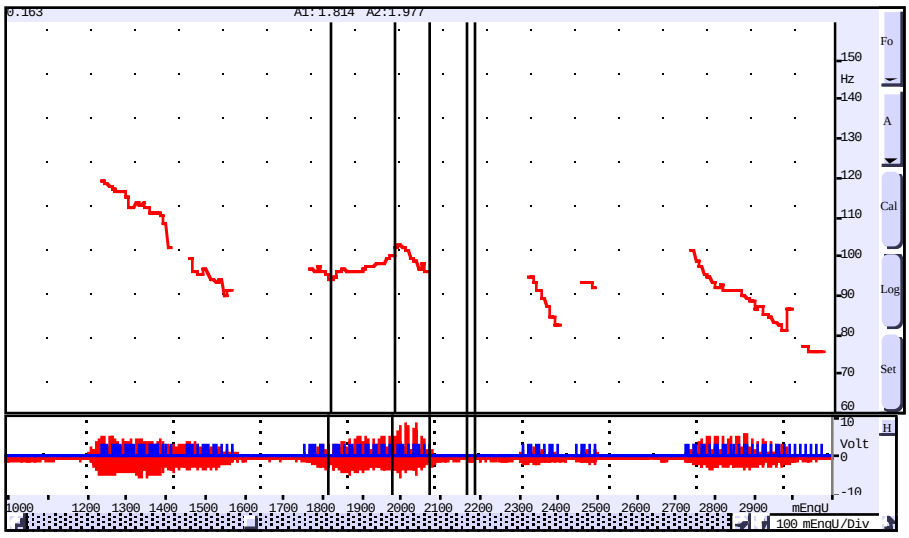

Figure 8 Wave plot and F_0 extract of "gild the lapses of time" in DF's reading.

The Empirical Background

The instrumental investigation of poetry reading has a long history; it goes back to the early twenties of this century. Much of this work was summarized by Schramm (1936). One cannot avoid therefore the question what distinguishes the present work from other empirical works. If one must give a one-sentence answer to this ques-

tion, it should be, I think, this: the present research conducts the instrumental investigations within the framework of a certain theory. This fact determines the significance of the answers found to a considerable extent. The experiments in this research were conducted to find answers to questions raised in the framework of a theory. Much of the early sound recorders' work was meant just to find out what happens in poetry reading. What is more, many of them were not exactly aware of a difference between the poem's structure and a casual performance. Some of them were measuring relationships between features of a casual performance, and thought they were obtaining information about the rhythm of the poem. A notable exception in this respect was Wilbur Schramm. Though he did not ask either sophisticated theoretical questions, much of his work can be regarded, in light of later linguistic theory, as contributions to the understanding of rhythmic competence. The present work is the first one, as far as I know, that establishes a *systematic* relationship between a poem's rhythmic structure and its rhythmical performance. The rhythmic structure is propounded on theoretical grounds; and the performance is conceived of as of a perceptual solution offered to a perceptual problem posed by the conflicting patterns of linguistic stress and metre. The process is constrained by the structure of the text on the one hand, and by Gestalt principles and the properties of short-term memory on the other. As so many times in my dealings with the semantic and phonetic aspects of poetic language, I had to discover that in this domain too Iván Fónagy (1959) preceded me, in noticing the "paradoxical" nature of poetic rhythm, and exploring some of the solutions offered by leading Hungarian actors to this paradox.

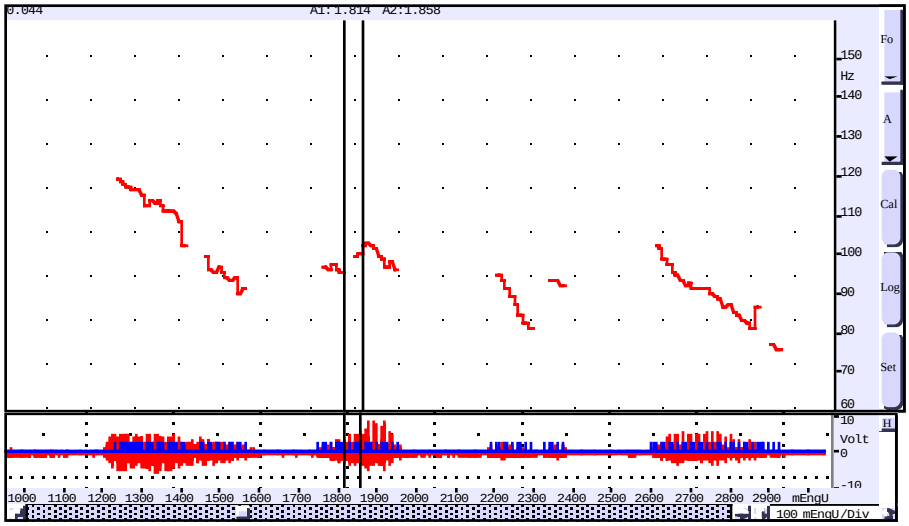

Figure 9 Wave plot and F_0 extract of "gild the lapses of time" in DF's reading after the removal of a 118 msec stretch from the /l/

As for the research tradition that approaches poetic rhythm from recordings of individual delivery instances, I shall indicate my position with reference to Chatman (1965), especially, and two younger scholars who are active now in Northern England. In an investigation of eleven commercially available recordings of Shakespeare's Sonnet XVIII by professional actors, Chatman (1965) used a panel of twenty one professors of English who were asked to make certain judgments. He went into enormous trouble to find out what the sonnet's metre was. He clipped segments of two consecutive syllables from each one of the eleven readings, which he played to his panel of twenty-one professors, and asked them to judge which one of them is more strongly stressed. As a result, he marked one atop the other sometimes as many as four different actualizations of one foot, as in line 2 (quote 8):

8. Thou art

He thought he was obtaining information about the poem's metre, whereas he was obtaining information only about the performances. Everything that can be known about the poem's metre, he could find out on his own, with a paper and a pencil, in his study. Any child could have told him that the possibilities marked in 8 were the four theoretical options. He didn't ask his panel of professors the most obvious questions, for instance, whether the greater prominence of *Rough* in one reading of "Rough winds" seemed to them acceptable. It is a breathtaking experience to read through this chapter of Chatman's. He gives some comparative information about the metric realization of the text by some of the greatest British and American actors. Then he gives some fascinating acoustic information about the ways they produced the stresses: e.g., "Marlowe Society's inflection on *Rough* is even steeper than Quayle's and the time-spans were almost equal; but the inflection on *winds* is sufficiently pronounced to overcome a trochaic bias. […] In Evans' recitation, however, extreme length of *winds* manages to overcome the very steep rise on *rough*, even though *winds* is practically level" (ibid., 174-175). I find all this stuff fascinating. This discussion of, e.g., "steep inflections" on *Rough*, might indicate, perhaps, how the performers utilize the acoustic cues for stress as information about the speaker's feelings toward rough winds; but it gives no new insight into the nature of stresses on the phonological level or, through it, into the solution of rhythmic problems. He did not identify problems to which these readings could serve as solutions; and did not formulate principles according to which these performances *could* serve solutions to problems. This may be called "multiplying information", without yielding insight or explaining observational facts on other metric levels. Had Chatman made a different use of the presence of his distinguished panel, for instance, had he asked them to judge which solution yields a better and which a less good rhythmic organization, he would have obtained, perhaps, a set of observational facts that could, perhaps, be explained through an analysis of acoustic cues. This is, of course,

less easily done than it sounds—I am merely suggesting ways in which the above kind of acoustic analysis could be used more fruitfully for a theory of metre. Listings of alternative possible actualizations of metric feet, or detailed descriptions of the acoustic information for their own sake tend to obscure issues: the structural problems to be solved by the performance, or the fact that the problems have *not* been solved.

The approach advocated here, by contrast, discusses acoustic cues or alternative realizations of metric feet in strict relation to their possible contribution to the solution of the problems posed by the conflicting patterns of the verse structure. An analogy might clarify this issue. I argued in my 1977 book that a distinction must be made between performance patterns and their acoustic and phonetic correlates. One of the performance patterns we have discussed and still will discuss at great length is "stress valley". The relationship between such a stress valley and its acoustic correlates is similar to the relationship between a phoneme and its acoustic correlates. We are interested in the phoneme as an abstract category, and ignore the specific acoustic cues that are its exponents. Consequently, there is usually a trade-off between the possible acoustic correlates that may cue a certain phoneme. Thus, for instance, a voiced stop may be cued by the straightforward activation of the vocal folds, or by a lengthening of the preceding sonorant, or by reducing voice-onset-time, or by aspiration. Most language users would not distinguish between the various vocal devices; they merely perceive a unified abstract category, such as [b, d, g]. It is the unitary performance category that counts, and not the indefinite variety of acoustic cues, among which considerable trade-off may take place, and even unforeseen combinations may be invented. The need for the performance category is determined, in turn, by the conflicting patterns of metre and linguistic stress. We shall see in the ensuing chapters time and again that performers are in fundamental agreement as for the rhythmical performance of some metric deviation (even those ruled unmetrical by two different criteria of generative metrics), by applying, as predicted by the present theory, e.g. a "stress valley" to the deviant segment; at the same time, the stress valley may be cued by a variety of acoustic and phonetic cues in the various delivery instances. This conception grants the reciter a great degree of freedom in resorting to acoustic cues, eliminating, at the same time, all notion of randomness or arbitrariness. I contend that there is an open list of possible acoustic cues. Reciters display an astonishing degree of creativity; new performances provide acoustic cues some of which are quite expected, but some are entirely unforeseen. But as long as they generate a unitary perceptual category and indicate the required segregation, listeners exposed to them for the first time immediately recognize them as appropriate (assuming that continuity is taken care of).

Chatma conceives of the term "ambiguous" in a prosodic context (both with reference to metric deviation and conflicting intonation contours) as "capable of alternative realizations at different times", rather than "reflecting conflicting structural principles at one and the same time". In utter disagreement with this conception, I have suggested (Tsur, 1977: 134 and passim) that where linguistic stress pattern or

intonation pattern conflict with the patterns required by versification, the performer may have recourse to conflicting cues. This applies, as the above examples show, both to enjambment and to instances where the linguistic stress pattern deviates from the metric pattern. Barney (1990) investigated the enjambments in two poems by Philip Larkin and John Betjeman read by the authors, and his findings confirm these predictions. The present approach however differs from Barney's in one respect of great theoretical importance.

Richard Cauldwell's paper submitted to *Language and Literature* (based on his doctoral dissertation) is primarily devoted to the empirical investigation of a different issue, namely, that "the procedures of interpretation of intonation choices [...] rest on the assumption that there is an intending speaker and a co-operative listener whose understanding is limited by their shared knowledge and shared expectations of what they are both 'up to'". His metrical notions are rather naive; but his following remark may pose a serious problem to Barney's procedure: "In Larkin's readings, all sense of the lineation disappears: the awareness of a ten syllable line and the position of rhyme words which comes from a visual consideration of the poem is undermined by the reading. Indeed this is characteristic of all the readings of 'Mr. Bleaney' and of his poetry readings in general". Indeed, Barney performs two very important stages of an investigation of enjambment, but not a third one: on the one hand, he performs a keen theoretical analysis of the enjambment instances, and establishes a scale of mounting complexities; on the other hand, he points out the conflicting cues for continuity and discontinuity in the authors' readings, and points out significant correlations between degrees of enjambment and cues for discontinuity. But he skipped the essential stage of listening by the ear to the auditory result. He took the authors' authority for a successful vocal performance of the enjambments. What Cauldwell's remark suggests is that in Larkin's case at least this should not be taken for granted at all. Curiously enough, in the present instance this seemed to work; but it should not be taken for granted. The present work meticulously checks the acoustic and phonetic cues against the perceived effect of larger segments.

Convergent and Divergent Delivery Styles

At the beginning of the present chapter I mentioned C.S. Lewis' distinction between the Minstrel, who subdues prose rhythm and foregrounds the metric pattern, and the Actor, who subdues the metric pattern in favour of the prose rhythm. I suggested that there is a third, "rhythmical performance", in which both metric pattern and linguistic stress pattern can be accommodated, such that both are established in the listener's perception. In the following section I am going to argue that within the legitimate boundaries of "rhythmical performance" metric pattern or prose rhythm may be more dominant. Accordingly, I shall speak of "convergent" and "divergent delivery styles" respectively.

Elizabeth Couper-Kuhlen writes in her book on English prosody:

> It is a basic principle of English speech rhythm that stressed and unstressed syllables alternate rather regularly. Consequently if an utterance contains a succession of, say, three monosyllabic words from stressable word categories, e.g. *'big 'black 'bugs*, the intermediate stress may be dropped in order to achieve a more regular alternation, e.g. *'big black 'bugs.* [...] Finally if two major stresses are adjacent to one another in a phrase or utterance, e.g. *thir'teen 'men*, under certain conditions one may be moved forward: *'thirteen 'men* [...] in order to more nearly approximate an alternating rhythm (Couper-Kuhlen, 1986: 37).

If instantiated in poetic rhythm, this might be a veritable prosodists' and performers' paradise: the number of deviant iambic lines would drastically decrease; that is, as long as the succession of three monosyllabic words ended in a strong position. If, however, the string began and ended in a weak position, as it sometimes happens, they were in a real trouble: in the sequence *'big black 'bugs* only the stress-demoted *black* would coincide with a strong position.[2] At any rate, in the above prosodists' and performers' paradise it is poetry that would suffer: most distinctions between e.g. Milton's and Pope's metrical style would be blurred.

How do, then, professional reciters handle such strings of successive stressed syllables? At the outset, one must point out that there are various delivery styles. Just as metrical style may be convergent or divergent, delivery styles too may be convergent or divergent (this is not a simple dichotomy with either/or choices, but rather a dichotomic spectrum with more/less distinctions). In convergent metrical styles, the metrical pattern and the linguistic stress pattern converge to a considerable extent; in divergent metrical styles they tend to diverge. Pope and Thomson would be typical representatives of the former style, Milton and Shelley of the latter. Shakespeare in his Sonnets is more divergent than Pope, but less than Milton (cf. supra, Chapter 1, figure 1). When divergences occur, the reciter attempts to accommodate the conflicting patterns in a third, superimposed, pattern of performance. Performers may attempt to minimize divergences, in which case they are said to have adopted a "convergent" delivery style; or they may emphasize divergences, adopting a "divergent" delivery style. Consider, for instance, excerpt 6 quoted above:

6. But thóu, contrácted to thine ówn bríght éyes, ...
   w s w s w s w s w s

In the last three syllables we have a succession of three linguistic stresses, whereas the metrical positions they occupy constitute a sequence of strong—weak—

[2] Some metrists have solved this problem in a rather elegant way: they speak of optional rules, which can be instantiated when it is convenient for them, and can be disregarded when inconvenient (cf. Beaver, 1971).

strong. The stress pattern is precisely of the kind discussed by Couper-Kuhlen. Even without dropping the stress entirely, the stress of a pre-nominal adjective is subordinated to the stress of the noun it qualifies. Thus, in convergent delivery style, the iambic lilt of the phrase may easily be preserved, without grossly violating the linguistic stress pattern. Nonetheless, some performers do not choose this "easy way out", and prefer, precisely, to emancipate "bright" from its subordination to "eyes", and perform the three consecutive words with equally heavy stresses.[3] Thus, for instance, Marlowe Society in their full recording of Shakespeare's Sonnets perform the words "own bright eyes" with three equally heavy stresses. The result is that the pitch pattern drastically deviates from what might be expected in spoken English. This would be a good example of divergent delivery style.

One possible reason for relinquishing here the facility available to the convergent delivery style is a wish to convey the metrical complexity of the verse line. Another reason may be a quest for consistency. The performer knows in advance that in the last line of the same sonnet he will face the sequence "To eat the wórld's dúe", where *wórld's* occurs in a strong position, whereas *dúe*, to which its stress is subordinated, occurs in a weak position. Here too the reciter effects a quickly rising and immediately falling pitch on *wórld's*, and a pitch falling to low on *dúe*. What is more, in the next sonnet the reciter is to face the phrases "thine ówn déep-súnken eyes" and "áll-éating shame". Each of these phrases contains a compound, in which the stress of the second noun is subordinated to the first one. Unfortunately for the reciter, the syllables *déep* and *áll* occur in weak position. Likewise, in the last two lines we have "This were to be néw-máde", where *néw* occurs in a weak position, and "And see thy blóod wárm", where *wárm* occurs in a weak position (these issues will be discussed at great length in Chapter 5). In anticipation of these and other strings of consecutive stressed syllables in which the linguistic stress pattern conflicts with the iambic pattern, divergent reciters may level the stresses from the beginning, even in those successions of stresses whose lilt might easily conform with the iambic lilt, attempting to create mental conditions in which the listener is capable of attending to his metrical set echoing in short-term memory. Convergent performers of such poems, by contrast, would seize any opportunity to convey the iambic lilt in the immediately observable constituents of language, wherever stress pattern can be made to conform with it, and attempt to find some acceptable, more or less convergent, solution to the rest as well.

Sometimes, one and the same reciter may resort to different delivery styles on different occasions in the performance of the same poem. Gielgud, for instance, in his

[3] This requires, as predicted by the present theory, heavy over-articulation of word boundaries. Couper-Kuhlen treats this as resulting from the constraints of articulation. As we shall see in chapter 8, what Couper-Kuhlen treats as a kind of "refrectory period", Thomas Cable mistakes, following some generative phonologists (Liberman and Prince; Selkirk) for observance of equal time periods in speech and poetry recital. He regards it as evidence for some "unrealised beat" in the versification pattern.

earlier recordings of Shakespeare, was inclined toward more convergent performances of consecutive stresses, whereas in his later recordings, he became more divergent.

Consider the following two lines, from Shakespeare's *Hamlet* and Sonnet 18, respectively.

9. Or to take arms against a sea of troubles
10. Nor shall death brag thou wanderst in his shade

Each one of these lines begins with two unstressed syllables, followed by two stressed ones. One obvious performance pattern for convergent performers would be what might be called a "stress slope": Suppose the performer stresses *shall* in excerpt 10 somewhat more strongly than *Nor;* and *Death,* in turn, somewhat more strongly than *shall,* and *brag* more strongly than *death.* Thus he would receive a group of four syllables which form part of one conspicuous shape that has a definite direction, without greatly distorting the stress patterns of language; at the same time, it would perfectly conform with the iambic lilt, in which the even-numbered syllables are stronger than the preceding odd-numbered ones.[4] Indeed, a similar stress slope is observed in Gielgud's reading of the above two lines in an early recording (on Brunswick LAT 8015). *Take* and *Death* are, in fact, not stressed. It is only the listener's grammatical knowledge rather than any acoustic correlates that allows him to perceive, in this performance, any stress on these two words. In a later recording of the Sonnets, Gielgud diminishes the difference between the stresses on *Death* and *brag,* whereas Marlowe Society lengthen the first four syllables of this line to a very great extent, and utter *Death* and *brag* with extra intonation on each syllable. In his later recordings Gielgud too offers a considerable number of equally strong consecutive stresses. Obviously, we are confronted here with two delivery styles, convergent and divergent.

Marlowe Society's performance of excerpt 6 is not without precedence in standard English prosody. Consider the following passage from Couper-Kuhlen:

> Several stresses *can* be articulated in immediate succession if the syllables bearing them are lengthened appropriately, or if pseudo-pause is inserted: *'big— 'black— 'bugs, thir'teen— 'men.* Likewise, additional stress beats need not be added in a sequence of unstressed syllables if the rate of delivery is appropriately fast. [...] The slower the rate of delivery, the closer together the stresses come (Couper-Kuhlen, 38).

This is, as we have seen, precisely the way in which Marlowe Society perform the equally strong consecutive stresses in excerpt 6. As for "Nor shall death brag" in

[4] Wimsatt and Beardsley propose such a performance pattern for similar lines (1958: 594). However, contrary to their general theoretical position, they do not realise that they are proposing *a possible* performance pattern, and not *the* performance of such lines (see below, Chapter 5).

10, Marlowe Society lengthen these syllables, that is, utter them more slowly than usual. Gielgud, by contrast, in his earlier recording speaks the sequences *take arms* and *Death brag* rather fast. Such a performance as Marlowe Society's has, as we have seen in Chapter 2, important implications for a cognitive account of the rhythmical performance of deviant lines. At the present stage of the argument, however, only one thing will be pointed out. Marlowe Society don't just improvize ad-hoc solutions: they *adopt a delivery style.* And that requires from them to adopt from the very beginning, e.g., a speech tempo or pitch contour which would allow them in due course to make the necessary deviations, without arousing a feeling of inconsistency. Such a delivery style does not presume to give an impression of *naturalness*, but rather an impression of *consistency*. The adoption of such a delivery *style* is what would be called *stylization.*

Now consider line 13 of Sonnet 12 in Gielgud's reading:

11. And nothing 'gainst Time's scythe can make defence
w s w s w s w s w s

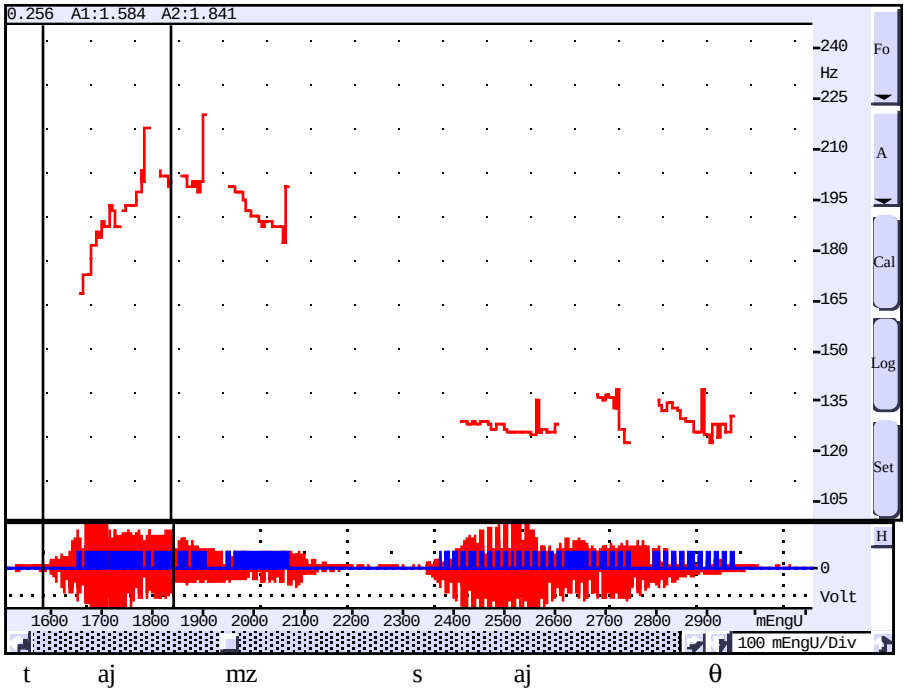

Figure 10 Wave plot and F_0 extract of "Time's scythe" in Gielgud's reading. The two markers indicate the boundaries of the diphthong.

Listen to sound file

In the phrase "Nothing 'gainst Time's scythe" we have a "stress grade": two unstressed syllables "thing 'gainst" followed by two stressed ones (both of which have a long diphthong for a nucleus): "Time's scythe". Such patterns are usually per-

ceived as rather unnatural in an iambic context. Suppose the performer stresses, in harmony with Wimsatt and Beardsley's commendation, *'gainst* somewhat more strongly than *thing;* and *Time's,* in turn, somewhat more strongly than *'gainst,* and *scythe* more strongly than *Time's.* Thus he would receive a "stress slope", a group of four syllables which form part of one conspicuous shape that has a definite direction, without greatly distorting the stress patterns of language; at the same time, it would perfectly conform with the iambic lilt, in which the even-numbered syllables are stronger than the preceding odd-numbered ones. Wimsatt and Beardsley claim this is *the* way to perform such verse lines; the present approach assumes that this is one possible "convergent" performance of such constructs.

The perception-oriented theory of metre predicts that, in a divergent delivery style, the performer would sharpen the contrast between the two unstressed and the two stressed syllables, so as to foreground the stress grade; whereas the two stressed syllables would bear, in discrepancy with the "nuclear stress rule", equal (heavy) stress. The iambic lilt would be present only as the reader's "metrical set", echoing in his short-term memory. In order to perceive both the metrical set and the immediately observable linguistic stress pattern at the same time, mental processing space must be spared by over-articulation of the phonemes and the syllable boundaries on the one hand and, on the other hand, by emphatic grouping of the stressed syllable in the weak position forward to the stressed syllable in the strong position. Consider the phrase "Time's scythe" in Gielgud's performance (figure 10).

The intonation contour of "Time's" steeply rises from 132.467 to 204.170 Hz; there are two vertical obtrusions at the end of the diphthong and on /m/, to 216.064 and 220.342 Hz respectively; then the curve descends to 182.414 Hz, rising again to 198.701 Hz. Even if we decide that the vertical obtrusions are artefacts, there is a very late humpback peak immediately before the diphthong boundary. The steeply rising and falling intonation contour takes care of the over-articulation of the syllable boundaries. The vigorous forward grouping is effected by two aspects of this intonation contour: according to the musicologists Cooper and Meyer, a steeply-rising pitch or amplitude contour typically effects forward grouping (it "leads", as it were, forward). And as I have pointed out time and again, late peaking generates an impetuous forward drive. In "Time's", we obviously have both a steeply rising intonation contour, and a very late peak (however, late peaking need not occur on a very high pitch).

Both nouns bear exceptionally strong stress, the former cued by steeply rising intonation, the latter by duration. The intonation contour of the second word is kept low, and right flat; its heavy stress is indicated by excessive duration: The second noun is considerably longer than the first one (0.728:0.906 sec). This ratio drastically increases if we consider only the duration of the nucleus, the diphthong /aj/ (0.256:0.469 sec).

The resulting effect is as follows: The heavily stressed syllable *Time's* (preceded by a function word occupying a strong position) hits the metric pattern in a weak position, and thus disturbs to a great extent metric regularity. The steeply-rising

intonation contour, and a very late peak arouse cravings for and expectations of what follows, grouping the noun with the next one. When *scythe* appears in the next strong position, it reaffirms metre arousing a feeling of relief and gratification.

To Sum Up

The instrumental investigation of poetic rhythm is, then, impossible *in principle.* One can only collect rhythmicality judgments from listeners, and look for vocal devices in the stretch of speech that typically count toward or against rhythmicality, *in light of the present theory.* The same set of vocal devices can account for the rhythmical solution of problems in cases of enjambment, strings of stressed syllables and stress maxima in weak positions—in this order of decreasing directness. Syntax in enjambment requires continuation, whereas the line ending requires stopping. This is obvious and straightforward. In case of strings of stressed syllables, the perception of the mental pattern of metre (metrical set) requires the over-articulation of word boundaries (discontinuation) in the immediately observable stretch of speech, whereas syntax demands in most instances continuity. The prosody of spoken English allows for the elimination of strings of equally stressed syllables; nonetheless, reciters who adopt a divergent delivery style prefer to preserve such strings, and over-articulate the word boundaries. The rhythmical performance of a stress maximum in the seventh (weak) position requires the segregation of the last four syllables in mid-phrase or in mid-word, and their foregrounding as a stress valley, while syntax, again, requires continuity. This is the most parsimonious way to save mental processing space for the simultaneous perception of the linguistic string and the mental set. Thus, in a variety of metric complexities, conflicting cues of continuation and discontinuation must be deployed by the reciter. The general pattern of simultaneous continuation and discontinuation is recurring; but the reciter may have recourse to a wide variety of specific cues. As will be seen in Chapter 6, performers are usually in fundamental agreement as for the general pattern, but display considerable creativity regarding the specific cues. Early and late peaking are additional vocal devices for the grouping of syllables. On closer inspection, they turn out to be instances of the wider Gestalt Principle according to which perceptual units tend to resist interruption, and when this happens, "perceptual forces" are generated. Such perceptual forces may be generated in enjambment too, or when a major syntactic boundary intrudes between the caesura and the line ending. In fact, as will be seen in Chapter 4, the caesura itself is conceived of as of a milder version of such an intruding event. As will be seen in Chapter 9, not as in music, where a pause replaces a note, in poetry recital pauses are perceived as such intruding events.