

Chapter 2

Rhythm

2.1. Introduction

In the foregoing chapter we have seen that language and music have similar kinds of structures and processes, both at the level of phrasing and at the rhythmic level. The remainder of this dissertation is about phrasing and rhythm phenomena in speech, while we keep musical phrasing and rhythm as the background motivation for several speech phenomena. The current chapter gives a (non-exhaustive) overview of the literature on rhythm, and provides the background information for both Chapter 3 on rhythm in fast speech, and Chapter 4 on recursive prosodic phrasing. We do not intend to contribute to the theoretical discussions involved in the rhythmic issues discussed in this chapter. The issues discussed are the following: in the first section we will describe what rhythm is. In the second section we will make the distinction between rhythm and meter. Section 2.3 shows that languages differ as to how their rhythm units are classified. Section 2.4 describes some variable rhythm patterns in speech, and the rules that seem to play a role in this variability. In section 2.5 we will point out that time is a major factor in rhythmic patterns, which also will turn out to be the core explanation for our observations in Chapter 3. Firstly, we will answer the question “what is rhythm?”.

Rhythm is everywhere, in the world within us and in the world around us. Rhythm is in our heartbeat, our breathing, and our stride, but also in the tides of the sea, the seasons and the movements of the earth itself. These are all movements in a rhythmic fashion. It turns out to be really hard to do things non-rhythmically: when people are asked to tap their fingers on the table irregularly, some recurrent pattern will appear (Fraisse 1982). This all suggests that rhythm is at the heart of nature, at least in natural events in which time,

movement, or visual patterns are involved. Language and music are just two of these rhythmical behaviors.

Rhythm usually implies some kind of succession of strong and weak elements of the same type, where the events occur with some repetitive structure, perceived at a constant rate, as a recurrent pulsation. Because the perception of a rhythm pattern needs some frame of reference to which the rhythmic parser can be set, it must be a series of at least three events or two (time) intervals to form a rhythmic pattern (Couper-Kuhlen 1993, Auer, Couper-Kuhlen and Müller 1999). A sequence of two events implies a single interval, which would not give the listener any reference to the internal relationships within the rhythmic pattern and the level on which regularity could be perceived. A sequence of three events gives us a succession of two (time) intervals, and this would give the listener an indication of the tempo and the regularity to be perceived.

Rhythm can be a pattern of any sequence, but mostly time and accentuation are involved (Randel 1986: 700-705). It can be thought of as the division of a temporal flow into perceptual groups. The grouping is what makes it rhythmical; without grouping the elements, it would be just a continuous flow of sound, going by unnoticed after a while. The elements are grouped by emphasizing some elements and by boundary marking. When we listen to sequences of pulses, any element that is louder, longer in duration, higher in pitch, or otherwise different from the rest, is perceived as leading the group of pulses. Combinations of differences in loudness, pitch and duration can lead to complex rhythmic groupings. The end of a rhythmic group is usually marked by a boundary marker, such as lengthening of the final element, deceleration of the tempo, or a pause after the final element.

Rhythm is only heard if a sequence of pulses is neither too rapid nor too slow (Woodrow 1951, Fraise 1963, Randel 1986). The minimum time between pulses of which successiveness and order are perceivable is about 0.1 second, whereas the maximum beyond which groupings do not form is about 3 seconds (Allen 1975). Phonological rhythms generally work at close range, within the phrase and not beyond the sentence (Allen 1975, Couper-Kuhlen 1993, Auer, Couper-Kuhlen and Müller 1999).

Perceiving rhythm is imposing some rhythmic structure on a sequence. Even where it does not exist in fact, rhythm is heard. A

nice example is the ticking of a clock, as already mentioned in the Introduction of this dissertation (Bolton 1894, Frawley 1963). The ticking is imitated by people by assigning the ticks different sounds, as if the two sounds differ. This indicates that we perceive the sounds differently, the tick sound emphasized with the high pitch of the intrinsically higher vowel /i/, the tock sound with a lower and therefore less prominent vowel /ɔ/. And by differentiating between the sounds, we automatically group them into groups of two, sometimes with a pause between the groups. This grouping behavior influences our perception of many temporally ordered entities.

Linguistic rhythm does not carry much information of its own, other than helping to guide the hearer's attention. Speech rhythm functions mainly to organize the information-bearing elements of the utterance into a coherent package, in order to make the informative elements temporally predictable in speech communication (*cf.* Chapter 3). Without rhythmic organization, the linguistic message would be difficult to transfer (Allen 1975).

2.2. Rhythm and meter

Rhythmic groups in speech and music can be regular or irregular patterns of different or similar elements. They consist of temporal relationships in a sequence of long and short sounds and silences, a free and creative ordering of time values (Broeckx 1967, Randel 1986).

Perfectly regular, recurrent rhythm constitutes meter (Randel 1986: 489, 702). Meter is the rhythmic level at which the conductor of an orchestra waves his baton and at which people dance. The term beat is sometimes used in a more abstract way for an arbitrary level of the metrical hierarchy. In music the term 'tactus' is used to indicate on which specific level the music is to be counted. It refers to the most salient periodicity or metrical level. This occurs between about 40 and 300 beats per minute, with a preference for a tempo of around 100 beats per minute, the so-called 'preferred rate' – a time interval of 600 ms (Frawley 1982). In the case of speech, this level is also called the level of scansion. Note that for popular music, music software such as Steinberg Music and Adobe Audition take 120 bpm as the default setting, and in techno music the default rate is even 140

bpm or higher (Noys 1995, Reynolds 1999). We can only speculate that the preferred rate, parallel to the tempo of society, is changing to a faster beat. On the other hand, this tempo difference could be a distinction between popular and classical music.⁶

Meter thus involves isochronic time intervals between rhythmically prominent elements. The question now arises how regular rhythm must be in order to be perceived as metric. Experimental literature about isochrony in music and speech fails to confirm its existence; intervals appear to be highly variable (cf. e.g. Bolinger 1981, Dasher and Bolinger 1982, Dauer 1983, Couper-Kuhlen 1993, Laver 1994, Terken and Hermes 2000, Fox 2000 among others for speech, e.g. Honing 2002, and Handel 1993 for music). Isochrony seems to be subjective rather than objective: listeners tend to expect intervals to be isochronic, and they tend to adapt their perception to their expectations. In order to perceive regular patterns, absolute precision is not required and considerable latitude is allowed without destroying the sense of isochrony. A tendency exists to underestimate the duration of long time intervals, whereas the duration of short ones may be overestimated (Allen 1975), with the result that shorter and longer intervals may be perceived as having equal durations (see also our results of Chapter 3).

Whereas rhythm is built on physical stimuli in the acoustical surface, such as differences in loudness, duration, or contour, meter is an idealization, a psychological percept of the rhythmic stimuli. Metric accent in music stems from the listener's active engagement with the music: his mind must perform a number of interpretive tasks to hear it (Couper-Kuhlen 1993, Auer, Couper-Kuhlen and Müller 1999). This can be shown if we vary the context in which a rhythmical sequence is perceived. In Figure 13 we see two melodies, where the only rhythmical difference between Figure 13a and Figure 13b lies in the first two measures. In the first two measures of Figure 13a the notes have a length of three counts, dividing the measure in two, which gives the impression of a bipartite meter with groups of three quavers, as in a 6/8 measure. Figure 13b on the other hand starts with three notes of two counts each measure, which causes the

⁶ Interestingly, hip hop music, which is a kind of (fast) speech, has a tempo in the same range as we find for speech, as described in Chapter 3: 80 to 120 bpm.

following quavers to be perceived as grouped in two by two, as in a 3/4 measure (London 2001).

Figure 13 Metric grouping



Meter is strongly predictive: metric patterns usually remain constant and are thus a reliable basis for anticipation when subsequent events will occur, while rhythm is not a reliable basis for anticipating subsequent events, although durational patterns may be repeated (Randel 1986: 489, Sachs 1953). Meter is also continuous, which means there are no gaps between successive groups. In the case of rhythm there are often gaps between successive groups.

The musical literature makes a clear distinction between rhythm and meter (Randel 1986: 702). In the linguistic literature, on the contrary, confusion of terminology arises, or one could just say that the terms are used differently. Mostly when phonologists speak of rhythm, they actually mean meter. Meter is seen as a subset of rhythm, namely as regular rhythm, which may in some cases be correct. The source of the confusion may lie in the fact that the perception of meter is often based on rhythmic events, and language, except for metric poetry, is usually not strictly regular. The distinction between rhythm and meter in language is therefore less easily made. The result of this terminological confusion can also be seen in Chapter 3 of this dissertation, where we first tried to find evidence for rhythmical shifts, but later found out that the perceived rhythms were only based on timing intervals, i.e., induced meter from signals differing in tempo. Although rhythm and meter are evidently distinct from each other, we will treat meter as the abstraction of regular rhythm in this dissertation, to fit in with the literature on linguistic rhythm, and also because the term 'meter' is already reserved for 'meter' as in prosodic foot structure (Kiparsky 1975), which differs subtly from the musical 'meter'. The reader should keep in mind the distinction we made above.

The perception of rhythm and meter in speech also depends on the structural levels of rhythmic organization. Languages differ in the levels on which rhythm is based. In the next section we will therefore introduce the typology of rhythm classes in languages.

2.3. Rhythm units

2.3.1. Rhythm typology

In linguistic rhythm a central role is played by the syllable as a structural unit. The syllable can be interpreted as a unit of length in its own right, apart from the length of its segments. The length of the segments are in fact determined by the syllable structure and syllable weight, as can be seen in e.g. lengthening in open syllables in Dutch. Syllables are combined into feet (Σ). The foot is fundamental in determining the positions of stressed vs. stressless syllables within words and larger strings. The structure of a foot can be characterized as consisting of a string of one relatively strong and any number of relatively weak syllables dominated by a single node (Nespor and Vogel 1986).

The different units formed by the prosodic hierarchy (Nespor and Vogel 1986) influence the rhythm patterns that are allowed. All levels of the prosodic hierarchy require their own rules of stress and accent assignment. Moreover, the boundaries of each prosodic level cause pre-boundary lengthening and pauses of relative strengths.

Traditionally, languages are classified according to three different rhythm classes: stress-timed languages, syllable-timed languages (Pike 1945), and mora-timed languages (Trubetskoy 1939: 171). The classification of languages into the different classes has long been controversial. More recently, however, more and more evidence has been found for the correctness of most of the classifications, on the basis of acoustic, perceptual and psycholinguistic investigations (among others Cutler et al. 1997, Ramus et al. 1999, 2003), although it is still not clear whether this classification into three groups is exhaustive, or that it should contain more categories, or that it rather should be seen as a continuous scale. Dutch and English are classified as stress-timed languages, French and Spanish are

exemplary for syllable-timing, and Japanese is the best-known example of a mora-timed language.

The acoustic properties of the different rhythm classes have been thoroughly investigated. Nevertheless, until recently most research has failed to confirm the existence of different types of isochronous intervals in spoken language (among others Bolinger 1965, Abercrombie 1967, Roach 1982, Dauer 1983). The impression is that in stress-timed languages two stressed syllables occur at approximately equal intervals of time, which implies isochrony between stressed syllables and rhythms of alternation. This would mean that the intervals between the stressed syllables in the title of this dissertation *pro'sodic processes in 'language and 'music* are equal. Stretches of unstressed words or syllables are therefore compressed, while adjacent stressed syllables are rhythmically separated by 'silent beats'.

In syllable-timed languages, on the other hand, one gets the impression that each syllable tends to be given the same space and that rhythm emerges because syllables (rather than stresses) occur at equal intervals of time. All syllables in a single phrase would take approximately the same time and make up rhythms of succession. Thus, in for instance *c'est absolument ridicule* 'it is absolutely ridiculous' all syllables would sound as isochronous.

In mora-timing the syllable would not be the basic rhythm unit, but the mora, the weight-bearing part of the syllable, is the unit of consistent length, and thus the rhythmical basis of the utterance. Morae consist of a short vowel and the preceding consonant. Some consonants, mostly nasals, can also serve as syllable nuclei on their own, representing one mora. Thus, the Japanese word *sensei* 'teacher' is quadrimoraic (*se-n-se-i*). In mora-timed languages the succession of elements of the same length is said to make up the feeling of a staccato rhythm (Fox 2000).

Dauer (1983) observed that for the division into stress-timed rhythm and syllable-timed rhythm some distinctive phonological properties play a role, mainly syllable structure and vowel reduction. Stress-timed languages have a greater variety of syllable types than syllable-timed languages and syllable weight plays a major role for stress assignment in these languages. With respect to vowel reduction, unstressed syllables in stress-timed languages usually have reduced vowels, are shorter or can in fact be absent. Recently, Ramus

et al. (1999) found that, besides these phonological properties, differences in vowel/consonant segmentation characterize the rhythm classes. These and other phonological and phonetic features combined with one another give the impression that some syllables are far more salient than others in stress-timed languages, and that all syllables tend to be equally salient in syllable-timed languages.

2.3.2. Musical rhythm parallels the linguistic rhythm typology

A number of musicologists and linguists have claimed that the prosody of a composer's native language can influence the structure of his or her instrumental music (e.g. Abraham 1974, Wenk 1987). However, this was never satisfactorily supported by experimental evidence. Patel and Daniele (2003) found a way to compare rhythmic patterns in the English and French languages and in classical music. They claim that spoken prosody leaves an imprint on the music of a culture.

Patel and Daniele (2003) used the speech materials and results from a linguistic study on rhythm structure by Ramus (2002). This study showed that British English, as a stress-timed language, had significantly higher values for variability of vocalic durations than French. The musical materials consisted of instrumental music of a relatively recent musical era, of composers who were native speakers of British English or French, who lived in England or France. The measurements were made on scores, not on performed music.

Their results show that there is much overlap between the English and French composers, but on average a robust difference emerges, which is in the same direction as the linguistic difference. The difference for music is smaller than that for speech, however, which is said to reflect within-culture variability. Nonetheless, they proved that the musical rhythmic differences of certain cultures parallel the rhythmic differences between native languages of those cultures.

In stress-timed languages such as Dutch and English, some rhythmic processes have been observed since decades. Rules like the 'rhythm rule' are common knowledge in the phonological literature on stress and accent (among others Liberman and Prince 1977, Hayes 1984, Selkirk 1984 among others). Before we turn to the empirical findings on rhythmic phenomena of this kind, the next section will first introduce the theories on the mechanisms which bring about

these phenomena. We will add some new data to these theories to show that more mechanisms seem to play a role and we will try to put these phenomena into explanations within the framework of Optimality Theory.

2.4. Variable speech rhythm

2.4.1. Eurhythmmy in speech

Rhythmic patterns, in speech as well as in music, can sometimes change due to several factors. In Chapter 1 we saw the effect of an asymmetrical principle of Eurhythmmy in language, the Phraseal Rule (PR). Hayes (1984) formulated two more, symmetrical, principles of eurhythmmy. Because Hayes' Eurhythmmy rules are set up for outputs, his theory can be seen as a kind of predecessor of OT. They can thus easily be used as OT constraints. The first of the Eurhythmmy rules is the Quadrisyllabic Rule (QR), which demands that a secondary accent in a phrase is ideally at a distance of four syllables from the main accent. For longer phrases it is the case that at the so-called 'level of scansion', the level immediately under the level on which the gridmark of the main accent is situated, the major beats are at a distance of two syllables from each other. In the traditionally cyclically derived structure of the phrase *Mississippi Déltá* this is not the case. The secondary accent on *sip* is just two syllables apart from the main accent on *Del* here. The QR thus prefers a secondary accent on the first syllable of the phrase and indeed the phrase is often pronounced as *Mississippi Déltá*. In this instance a conflict arises between the QR and a correspondence rule which, on the basis of *Mississíppi*, prefers the accent to fall on the penultimate syllable. We will start from the assumption that different rhythmic structures imply different constraint rankings, following Schreuder and Gilbers (2004b). In fast speech the QR is dominant. Note, however, that we will arrive at an alternative approach towards the end of this chapter, which will be elaborated in Chapters 3 and 4.

Table 10 OT analysis

constraints → <i>Mississippi Delta</i> candidates ↓	QR	CORR
☞ <i>Mississippi Déltà</i>		*
<i>Mississippi Déltà</i>	*!	

The QR divides language into a kind of 4/4 measure. In tonal music this is also a common kind of measure, in which – as was said in the foregoing section – the first count in the measure is important and the third count receives a lighter accent. This last effect is seen again in the Eurhythmy Rule known as the Disyllabic Rule: immediately under the level on which the major beats are at a distance of four syllables from each other, the major beats are ideally two syllables apart. A rhythmic pattern satisfying these Eurhythmy rules shows a regular alternation of strong and weak elements at all levels (*cf.* TSRPR 1, Chapter 1). In Table 8 we see the interaction of the different Eurhythmy Rules with the OT constraint CORRESPONDENCE.

Table 11 Eurhythmy Rules in *twenty-seven Mississippi legislators*

constraints → <i>twenty-seven Mississippi legislators</i> candidates ↓	PR	QR	DR	CORR
X X X X X X X X X <i>twenty-seven Mississippi legislators</i>	*!	*		
X X X X X X X X <i>twenty-seven Mississippi legislators</i>	*!			**
X X X X X X X X <i>twenty-seven Mississippi legislators</i>		*!*		*
X X X X X X X X <i>twenty-seven Mississippi legislators</i>				**

In the first output candidate both the PR and the QR are violated, and it is therefore rejected as an optimal candidate. The second candidate is also rejected on the basis of the PR. It violates CORR twice, but it was out of competition already because of the fatal violation of the PR. The third candidate is fine with regard to the PR, yet it violates the QR twice: once between the main stress on *le* and the syllable *sip*, with a distance of only two syllables, and once between *sip* and *twen*, which has an overlong interval of six syllables. The first violation of the QR is already fatal, because a better candidate, which satisfies the QR, can be found. Candidate 4 is the optimal candidate of the candidate set in this constraint ranking, in spite of its two violations of CORR. All three – higher-ranked – eurhythmy rules are satisfied by this output candidate.

These rhythmic preferences in speech show that the rhythmical structure may change. The question is what kind of properties influence these changes. In Chapter 3 we try to answer the question whether speech rate can lead to rhythmic restructuring. In the next section we will first look at some other varieties of variable linguistic rhythm.

2.4.2. Triplet rhythm in trochaic Dutch

Besides the eurhythmy rules, a very well-known phenomenon in speech is clash avoidance (Prince 1983): clashes of strong, stressed, or accented syllables are avoided to prevent the rhythm from becoming ‘staccato’. This phenomenon is formulated in the constraint *CLASH. A preference seems to exist for beats that are more evenly distributed over the phrase. In Table 12 we see that the distribution of strong syllables over the word *bijstanduitkeringsgerechtigde* ‘person entitled to social security’ in an andante tempo is different from the distribution in an allegro tempo. The phrase gets a triplet-like rhythm at allegro tempo. Gilbers (1987) notices that in fast speech the fourth syllable is stressed more than the third. In andante speech, however, the first syllable in *uitkering* gets more stress than the second.

Table 12 Rhythmic structure *bijstanduitkeringsgerechtigde* (Gilbers 1987)

a. andante	<i>bij stand uit ke rings ge rech tig de</i>
	s w s s w w S w w
b. allegro	<i>bij stand uit ke rings ge rech tig de</i>
	s w w s w w S w w

Neijt and Zonneveld (1982) and Van Zonneveld (1983) have argued that Dutch is a trochaic language. In OT terms this means that the constraints RHYTHMTYPE=TROCHAIC (RT=T), a foot consists of a strong syllable followed by a weak one, and FOOTBINARITY (FTBIN), a foot consists of two syllables (or two morae), are high in the

constraint ranking for Dutch. The question now is how these constraints account for the triplet rhythm in Table 12b.

Gilbers and Jansen (1996) formulate an elaborate OT grammar for Dutch stress patterns, partly based on Nouveau (1994), Van Oostendorp (1995) and Kager (1994). The relevant part for this chapter is given in Table 13.

Table 13 Ranking for rhythmic base structure in Dutch

RT=TR ; FTBIN >> PARSE σ >> ALIGN-PRWD >> ALIGN- Σ

This grammar enables us to describe the longer rhythmic patterns in Figure 14.⁷

Figure 14 Possible rhythmic patterns in Dutch for phrases with more than four σ 's

- a. (σ σ) σ (σ σ)
- b. (σ σ) (σ σ) (σ σ)
- c. (σ σ) σ (σ σ) (σ σ)
- d. (σ σ) (σ σ) (σ σ) (σ σ)
- e. (σ σ) σ (σ σ) (σ σ) (σ σ)

In the OT grammar PARSE- σ demands that syllables are parsed in a foot and FTBIN and RT=TR provide for the preference of trochaic feet. These constraints dominate PARSE- σ , which results in an unparsed syllable for every phrase of an uneven number of syllables. ALIGN- Σ demands that feet align with the right edge of the phrase. This constraint, however, would cause the unparsed syllable with an uneven number of syllables to align with the left edge of the phrase. This is avoided by a dominant alignment constraint ALIGN-PRWD which requires that the left edge of a phrase aligns with the left edge of a foot: ALIGN (PrWd, Left, Foot, Left). This constraint now ensures that a triplet pattern can only occur at the left edge of a phrase.

⁷ For the sake of clearness we abstract from the influence of syllable weight here, and thus we start from a sequence of light syllables.

The ranking in Table 13 implies that the grammar does not enable us to describe the triplet pattern in Table 12b, because in that example the triplet pattern is carried through the whole phrase. The OT grammar will always prefer the structure in Figure 14e to that in Table 12b for a nine-syllable phrase. After all, the constraint ranking in Gilbers and Jansen (1996) leads where possible to a trochaic rhythmical basic pattern for Dutch phrases, like in *parallelogrammen* ‘parallelograms’, where s- and w-syllables nicely alternate. A compound like *bijstanduitkeringsgerechtigde* also gets a trochaic pattern – where allowed by morphological structure and difference in syllable weight (as in Table 12a).

However, there are languages in which the standard is a ternary rhythmic pattern. The dactyl pattern in Table 12b is a normal rhythmic pattern of prosodic words in languages like Estonian and Cayuvava: every s-syllable alternates with two w-syllables. Kager (1994) bases his analysis for patterns like those on *Weak Local Parsing theory* (Kager 1993; Hayes 1995): feet are at most binary and a ternary pattern is caused by an unparsed syllable between feet.⁸ The constraint bringing about this effect is FOOT REPULSION: *ΣΣ (avoid adjacent feet), it is a kind of OCP effect. FOOT REPULSION dominates PARSE-σ in languages like Estonian and Cayuvava. This constraint allows us to account for the alternative rhythmic structure of *bijstanduitkeringsgerechtigde* in Table 12b.

The basic assumption in OT is that constraints are universal. Because of the trochaic character of Dutch we have to assume that FOOT REPULSION is situated very low in the constraint ranking of Dutch, in any case it is dominated by FTBIN and PARSE-σ. In fast speech such a constraint ranking leads to a pile-up of many accents in a short period. Therefore, a longer interval between the accents is needed. In order to avoid clashes, beats are distributed over the phrase as evenly as possible. The distances between beats are enlarged, in order to avoid a staccato-like rhythm. Evidently, we are not dealing with phonetic compression here, but with restructuring, which can be obtained by assuming a special constraint ranking for fast speech (see Chapter 3 for a different analysis, however). In fast

⁸ Contrary to the analysis of ternary patterns like (s w) <w>, Dresher and Lahiri (1991), Selkirk (1980), and Hewitt (1992) propose ternary feet: (s w w). Dresher and Lahiri propose an extra parameter, in order to achieve a branching head of a binary foot. The resulting ternary feet violate FTBIN.

speech FOOT REPULSION (*ΣΣ) dominates PARSE-σ. The different constraint rankings for andante and allegro speech will then be as in Table 14.

Table 14 Rhythmic variability and speech rate

- a. andante ranking:
Rt=Tr ; FtBin >> Parse-σ >> Align-PrWd >> Align-Σ >> *ΣΣ
- b. allegro ranking:
Rt=Tr ; FtBin >> *ΣΣ >> Parse-σ >> Align-PrWd >> Align-Σ

As a general rule we will hypothesize in Chapter 3 that in allegro rankings markedness constraints and OCP-effects (*CLASH, *ΣΣ) are dominant, while in andante rankings CORRESPONDENCE constraints are far more important. Consider the different pronunciations of the word *tandpasta* ‘tooth paste’ in Table 15. In addition to clash avoidance there is a preference for assimilation in allegro styles. The functional explanation is that markedness constraints (ease of articulation) dominate correspondence constraints (ease of perception) in fast speech. Speaking fluently becomes more important and therefore unmarked structures are preferred from an articulatory point of view.

Table 15 Andante and allegro speech

	a. andante:	b. allegro:	
<i>tandpasta</i>	[tanpasta]	[tampəsta]	‘tooth paste’
	S s w	S w s	

Besides these kinds of rhythmic restructuring, we already spoke of the Phrasal Rule in Chapter 1. The preference for satisfaction of boundary-marking constraints (HAMMOCK) at the expense of violation of correspondence constraints (O-O CORR with the base of the adjective *tandheelkundige* ‘dentistry’) can also be seen in the data

Most rhythmic adjustment phenomena we described in this section have not been based on systematic, objective testing. Cooper and Eady (1986) tested the eurhythmy rules empirically and they did not find systematic evidence for the stress adjustments reported in the literature. The observations of such phenomena are persistent, however, and the question is whether we should view these phenomena from a different perspective. What catches the eye is that all these alleged stress phenomena aim at regularity and rhythmic alternation. A different perspective could be that this regularity is not based on rhythm as syllable counting, but on rhythm as timing, or, following the section on rhythm and meter in this chapter, on the more abstract notion of meter. There seems to be some mechanism that requires stresses to fall on an ideal time interval from each other. This points to a dominant constraint 'METRONOME'. In the next section we will describe some more recent accounts of rhythm, on music as well as on speech, that take rhythmic timing into account. Chapter 3 will also lead us to the conclusion that rhythm perception is based on timing.

2.5. Rhythmic timing

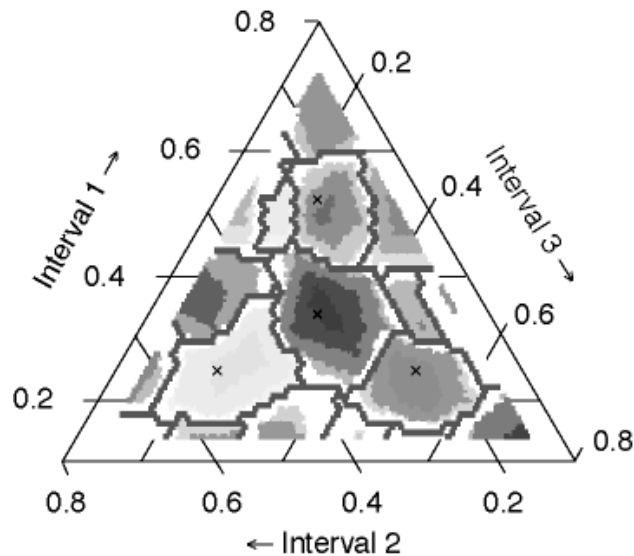
2.5.1. Restructured rhythm in music

A question in musical rhythm research is whether tempo changes imply relative invariance, or considerable deviation from proportional scaling. In other words, do the time intervals stay in the same ratio relative to each other when tempo changes, or are the relationships between rhythmic intervals distorted by the changed rate? Will all notes keep the same relative durations? Or will some notes get more prominence at the cost of other notes? This is a quite similar discussion to the one we will pick up in Chapter 3, where the notion of phonetic compression is the same as relative invariance and proportional scaling here, and rhythmic restructuring as the deviations from them.

Repp (1994, 1995) finds that in expressive timing, which expresses the phrasal structure of the performed music, only small deviations from proportionality occur. However, on the lower levels of rhythmic timing, considerable deviations from proportional

scaling are reported (Desain and Honing 1994). At different tempos, rhythm patterns are performed and perceived differently. Listeners do not perceive durations on a continuous scale. Instead, they recognize rhythmic categories that function as a reference relative to which the deviations in timing can be perceived. At faster rates, fewer categories of rhythm are perceived and more complex rhythms tend to drift to more simple rhythms. Honing (2002) and Desain and Honing (2003) define the simple rhythms and the related but more complex variants ‘surrounding’ them as ‘rhythm spaces’, as shown in Figure 16. The rhythm spaces are the categories in which in the listeners’ perception different rhythms tend to shift to the central rhythm. This is comparable to the way allophones are perceived as phonemes. Though all the allophones may be different, they group around the ‘central’ phoneme, and they are all perceived as one phoneme in the specific language.

Figure 16 Ternary plot of the rhythm space showing the areas of responses, centered around the points of the different rhythmic stimuli (Desain and Honing 2003)



Povel (1981) already found that subjects' reproductions of musical rhythms, although related to the stimulus ratios, are strongly distorted in the direction of a ratio of 1:2, 1:2 itself being reproduced most accurately. Collier and Wright (1995) also observe tendencies of reduction towards simple ratios at faster tempos, and towards contrast at slower tempos. The simple ratios to which the rhythms are drifted are never actually achieved, however (Repp 1998, Repp et al. 2002); the deviations from the optimal ratios probably lie within the same region of the rhythm space (Honing 2002). The preference for simple ratios implies that rhythmic patterns with many different note durations will be equalized to some extent in production, and are also perceived as more equal than they are.

The deviations of the rhythm are not proportionally scaled when tempo changes, at least in larger rhythmic groups. Repp et al. (2002) show that for three-note rhythms short intervals are timed quite accurately, whereas longer intervals show assimilation which increases as the tempo increases. Desain and Honing (1993, 1994) and Repp et al. (2002) conclude that relational invariance does not hold for rhythm at varying tempos: rhythm patterns are restructured. Tempo changes lead to a perceptual reorganization or regrouping of the rhythmic structure; long intervals turn into short intervals, and larger groups of events are formed. Sometimes it is even difficult to recognize the same rhythm at different tempos (Handel 1993). All results mentioned above point to the fact that rhythm is variable, and that it is strongly related to timing. This finding in music is the basis for our analyses in Chapter 3. In the next section we describe some accounts of speech rhythm as a timing phenomenon.

2.5.2. Timing in speech rhythm

In the perception of speech rhythm, not only the alternation of prominent and less prominent events are at stake, timing also plays a primary role, as we saw for music in the foregoing section. Port et al. (1998), Cummins (1997), Cummins and Port (1998), and Quené and Port (2003, 2005) integrate a mechanism known from extra-linguistic physical dynamical systems, self-entrainment (*cf.* Kelso 1995), into speech science. Self-entrainment means that the timing of repetitive motions by one oscillating system, like the pendulum of a clock, influences the motions of the other oscillator such that they fall into

simple temporal relationship with each other. That is, the oscillators tend to perform their motions in the same amount of time or in half (or double) the time, or in some other simple integer ratio of time. Speech rhythm is found to resemble such oscillating systems (Cummins 1997).

Cummins (1997) performed a couple of experiments with speech cycling tasks, and he found that the effect of harmonic timing is highly robust under conditions that encourage settling, not only in speech cycling tasks (repeating the same phrase), but also in chanting, poetry reading, singing while working together, etc. Speakers have a strong tendency to place the onsets of stressed syllables at temporal harmonic fractions of a metronome cycle, preferably at the largest harmonic fractions of 1: 1/2, 1/3, 2/3. These cycles are interpreted as isochronous series, if series can include silent beats as well (*cf.* Abercrombie 1965). Notice that this replicates the findings for musical rhythm (Povel 1981, Collier and Wright 1995, Honing 2002, Repp et al. 2002).

Quené and Port (2003) performed an experiment in which they aligned syllables of short phrases, such as *big for a duck*, with a metronome beat. They show that prominent vowel onsets are attracted to periodically spaced temporal locations, and this finding was formulated in the Equal Spacing Constraint. Vowel onsets are seen as the best approximation of P-centers (perceptual centers), the reference points for syllables on the basis of which listeners judge the timing of syllable or word sequences (Morton et al. 1976, Patel et al. 1999, Pompino-Marshall 1991).

Quené and Port also show that the Equal Spacing constraint gets stronger when the speaking rate increases, with an increasing number of stress shifts at faster speaking rates as a result. In fast speech, the number of beats is reduced, which appears to lead to a different prosodic structure containing less stressable positions, i.e. fewer feet. These feet in fast speech can consist of two or three syllables, or optionally of a single syllable supplemented with a silent beat: [*big for a*] [*duck* _]. This finding suggests that produced stress shift is conditioned at least in part by the global rhythmic pattern of the utterance. Quené and Port (2003) performed some reaction time experiments on spoken word perception with more natural speech data. They find that rhythmical regularity helps in word recognition.

These findings are highly relevant for our investigations in secondary stress shifts and rhythmic restructuring in fast speech, therefore we will also take up this line of investigation in Chapter 3. The experiments in the papers discussed above, except for Quené and Port (2005), use quite unnatural speech tasks. In our own experiments we tried to find our evidence in a more natural experimental setup in order to get near-natural speech data. In the next chapter, we will test our hypothesis that speech rate, just as in music, influences the rhythmic structure.

2.6. Summary

In this chapter we gave an overview of some issues concerning rhythm in music and speech. In the first section we saw that rhythm is quite an omnipresent phenomenon in the world around us, which means it is not unique to speech or music. This chapter is concerned with rhythmic sound, i.e. patterned sound. We illustrated that rhythm can consist of either regular or irregular patterns, grouped into structures using accents to mark prominent elements.

The second section distinguished rhythm and meter. While rhythm consists of physical stimuli, meter – perfectly regular rhythm – is merely a psychological, subjective percept.

In the third section we showed that the linguistic rhythm-bearing units are the mora, the syllable, and the foot, and that rhythm is also influenced by stress patterns. On the basis of these rhythm units, a classification into three rhythm classes has been drawn up, into which languages of the world have been classified. Moreover, a nice parallel has been found between the musical rhythm of certain cultures and the rhythm classes of their languages.

In section 2.4 we described variability in speech rhythm, and we saw that speech rhythm has some ideal patterns to which it tends to conform. Fast speech often triggers triplet patterns. We introduced the constraint $*\Sigma\Sigma$ (Kager 1994) in the constraint ranking to account for these triplet patterns.

In the last paragraph we discussed what happens to rhythm patterns under the influence of tempo changes. In speech as well as in music, the patterns tend to shift towards simpler ratios.

The last two subjects are the basis of the next chapter, in which we will investigate rhythmic changes under the influence of speech rate.