Chapter 3

Phonetic and Phonological Aspects of Tempo Variation

Introduction

What happens when we talk faster than we normally do? And what happens when we talk slower than we normally do?

There are many phenomena observable in speeded up and slowed down speech, which can be assigned to different levels of speech production. It involves "high" levels of speech production such as the prosodic phrasing as well as "lower" levels such as the velocity with which articulator movements are executed during speech production.

This chapter aims to give a comprehensive view over the levels and the mechanisms on each level which might operate while speaking at tempos different to the "normal" one.

3.1. Phrasing and pausing

As pointed out in the previous chapter, it seems generally accepted since Goldman-Eisler (1968) that changes in tempo are mainly changes in pausing rather than changes in articulation rate. To vary the tempo, speakers change the number of pauses and change the duration of pauses. Consequently, slowing down is characterised by more pauses and longer pauses compared to speech at normal speed, and speeding up features fewer pauses and shorter pauses compared to normal speeded speech.

Indeed, Caspers & Van Heuven (1991) found deletions of phrase boundaries in fast Dutch speech. In addition to boundary deletions, demotions of phrases from major to minor phrases were reported for two of three French subjects in Fougeron & Jun (1998). A perception test with Dutch subjects revealed that sentences with an

intonational phrase boundary in the contour were perceived ceteris paribus as slower (Rietveld & Gussenhoven, 1987).

In studies of French material read at different rates (Grosjean, 1979; Fletcher, 1987; Fougeron & Jun, 1998) the following observations have been made: slowing down is characterised by increasing, speeding up by decreasing the number of pauses. The higher the speed, the fewer the pauses. Evidence in (Bartkova, 1991). For German, however, Butcher (1981) observed different patterns for speeding up and slowing down: Increased tempo was characterised by reduced pause duration without a change in the number of pauses; slowing down was marked by a greater number of pauses without a change in the mean pause duration.

No clear picture about the reduction of pause duration arises from the French investigations. Fougeron & Jun (1998) found a reduction for their three speakers, only three out of six speakers applied pause shortening in Fletcher (1987), and Grosjean (1979) reported relatively constant pause durations. This is in strong contrast to Minifie (1963, cited in Lass, 1971) who, prior to Goldman Eisler, sees tempo variation more as a function of the compression and expansion of pause intervals than as a variation in articulatory intervals.

Table 3.1. increase (>), decrease (<), or no difference (=) to normal speeded tempo regarding number (no.) and duration (dur.) of pauses.

	slowing down		**speeding up**	
study	**number**	**duration**	**number**	**duration**
expectation	>	>	<	<
Grosjean (1979)	>	=	<	=
Fletcher (1987)	>	?	<	(<)
Fougeron & Jun (1998)			<	<
Butcher (1981)	>	=	=	<
Minifie (1963)	=	>	=	<

A naïve expectation might be that most speakers make maximal use of pausing mechanisms, namely reducing the number and duration of pauses for speeding up, and

increasing the number and duration of pauses for slowing down. From this brief analysis of pausing strategies across different studies and different languages, it becomes clear that this is obviously not the case, and that different speakers make use of the various pausing mechanisms differently. Table 3.1 summarises all the main findings from the above mentioned studies. It would appear that there is *not* a tendency to use *all* possible pausing mechanisms to change tempo.

But despite these restrictions on the full use of pauses as a means of tempo variation, it is clear from table 2.1 (in the previous chapter) that the relative amount of pause time can be up to half the entire speaking time. The central role that pausing plays in the regulation of tempo requires a detailed answer to the question: what is a pause?

Definition of pauses

A first distinction must be made between a *perceived pause*, and an *acoustic pause* or a *silent pause*: listeners, when asked to react as they hear a pause, tend to miss some acoustic pauses at unexpected locations. Likewise they interpret pauses at grammatical locations where, actually, there was no acoustic silence but only other phrase boundary markers, such as phrase-final lengthened syllables and intonational markers.

However, an acoustic pause is by no means always silent. *Breath pauses*, characterised by an inhalation noise, belong to acoustic pauses. A further distinction has to be made between *filled pauses* and *unfilled pauses*. Filled pauses usually occur in spontaneous speech; e.g. hesitation phenomena caused by planning problems or as discourse markers in dialogues. Although there is a debate whether fillers are words or not (cf. Clark & Fox Tree, 2002), there seems to be agreement on the phonetic content of the fillers. Typically, fillers are either [ɛː] or [ɛmː] or [ə]. In contrast to filled pauses, unfilled pauses usually occur at grammatically motivated locations and usually consist of silence. However, some unfilled pauses are phonetically filled, namely with breathing. Of course, all combinations of breathing, fillers and silence can be found.

For an exact definition of a pause, it is necessary to determine a threshold where a pause starts to be a pause. While some studies work with a threshold below 50 ms (pause and segment detection with an automatic procedure) (e.g. Lee & Oh, 1999), in other (mainly psycholinguistic) studies only silent pauses longer than 500 ms are

regarded (for a review see O'Connell & Kowal, 1983). Other examples are 200 ms (Grosjean & Collins, 1979); 150 ms (Tsao & Weismer, 1997); 130 ms (Dankovičová, 1997); 100 ms (de Pijper & Sandermann, 1994). Other silent intervals can complicate the pause defintion. The closure phase of plosives as articulatory activity should, of course, not be considered as a pause, although it is a silent interval. The closure phases typically ranges between 30 and 70 ms. But problems arise when an inter-pause stretch starts with a plosive. There, it is usually impossible to find the acoustic reflex of the beginning of the closure in the waveform. Then, either a default value must be set for all stops (or different classes of stops, e.g. fortis vs. lenis) or, as a less accurate method, these closure phases have to be ignored.

Realising prosodic boundaries

From a phonological point of view, pauses can be considered as phonetic realisations of prosodic boundaries (or prosodic breaks). Related to the claim that pausing plays the central role in speech tempo variation, it could be assumed that prosodic phrase structure, with pauses as the main markers of the prosodic boundaries, plays a central role in speech tempo variation. Therefore, further markers responsible for prosodic breaks should also be considered such as:

- the presence of an intonational boundary tone expressed as a F0 contour (e.g. de Pijper & Sanderman, 1994)
- phrase-final lengthening of specific parts of segments (e.g. Oller, 1973; Strangert, 1991; Turk, 1999)
- change of voice quality, e.g. creaky voice & whispering (e.g. Strangert, 1991; Kohler, 2000)
- declination in intensity and spectral tilt (e.g. Streeter, 1978; Strangert, 1991; Trouvain et al., 1998)
- phrase-initial strengthening (e.g. Fougeron & Keating, 1997)
- phrase-initial glottal reflexes, e.g. glottal stop and laryngealisation (e.g. Rodgers, 2000)

In the case of the wide-spread falling terminal intonation contours, all the above listed parameters fit into a picture where the acoustic silence is preceded by an utterance-final relaxation (Kohler, 2000) and followed by an utterance-initial tension.

However, for marking prosodic boundaries, all the above listed parameters are not necessarily fully realised and some of them may be missing. It seems generally accepted that silent intervals, F0 resets and finally lengthened sound segments are the primary cues which signal a boundary (Sanderman & Collier, 1996), whereas all the other parameters are considered secondary cues (Lehiste, 1970).

Levels of pauses and levels of prosodic phrase boundaries

Usually, there is more than just one level of prosodic phrase boundary. In many studies, a three-way distinction is made in major break, minor break, and no break (e.g. Crystal, 1969). The ToBI (Tone and Break Indices) scheme assigns six levels of breaks, with …

Intonation phrase normally marked by a pause …

There were also attempts to classify pauses into distinct categories. A division in *four* categories of pause length has been proposed by different authors such as Crystal (1969) (brief – unit – double – treble pauses). Butcher (1981) mentioned long pauses with corresponding 1400 ms, short pauses (500 ms) and unheard pauses (150 ms). In the study by Bartkova (1991) short pauses were those below 200 ms, medium ones were between 200 and 500 ms, and long ones were those exceeding 500 ms.

It can be implicitly assumed that a higher level break is marked by a longer pause than a lower level break. This is also what Strangert (1991) claimed for her Swedish news-reading data. She found a positive correlation between the acoustic signalling and the rank of the boundary for pauses as well as for other break markers such as F0 resetting and voice quality irregularities, but not for phrase-final lengthening. That means in relation to tempo, e.g. for fast speech, that the typically shorter durations are either a consequence of a general shortening of the pause durations, or that the breaks are stepped back to a lower level, or that both mechanisms operate.

Placement of pauses and prosodic boundaries

So far, we have learned that tempo variation can be reflected in a change of the number of pauses and a change of the duration of pauses. But an essential question is

still open, notably which pause, i.e. at which location, should be skipped or inserted, or shortened or lengthened, respectively.

Lass & Deem (1971) found an increase in the number of within-sentence pauses for slow readings whereas fast readings were characterised by fewer between-sentence pauses.

Bartkova (1991) shows with a three-way division of pauses, that in her texts the number of pauses at „full stops“ are relatively constant over three rates; at „commas“, pauses change slightly (but more in the fast direction), and syntactic pauses show the greatest variability.

Strangert (1991) showed for her Swedish data that pauses at paragraph boundaries are by far the longest, followed by pauses at sentence boundaries, and with pauses at clause and syntactic phrase boundaries the shortest.

Caspers & van Heuven (1991) found that pauses at obligatory phrase boundaries do not tend to be skipped at fast rates, but those at voluntary phrase boundaries do.

In an investigation of German news reading, Mixdorff (2002) found that pauses at major phrase boundaries between sentences were considerably longer than major break pauses within sentences. Additionally, a noticeable number of the within-sentence pauses were detected at locations *not* marked by punctuation in the text. This, of course, has serious implications for pause prediction, and therefore tempo modelling in text-to-speech synthesis.

3.2. Intonation

There are several studies dealing with the effect of tempo on intonation, such as number of pitch accents, the choice of accent types and the truncation and compression of accentual F0 movements, pitch excursion size and overall pitch range, and the temporal alignment of pitch accents.

Caspers & van Heuven (1991) found no difference between normal and fast speech concerning the number of pitch accents.

Dutch listeners in Rietveld & Gussenhoven (1987) were asked to rate sentences that differed only in the complexity of the intonation contour. Examples with complex structures were judged ceteris paribus as slower than less complex structures. Correspondingly, in the Dutch production data in Caspers & Van Heuven (1991) pitch

configurations marking the right boundary of intonation phrases tended to be simplified in fast speech. Similarly, there are simplifications of pre-nuclear pitch patterns. Fougeron & Jun (1998) e.g., report deletions of the initial high tones in French.

Ladd et al. (1999) found in their English data that F0 excursions are larger as rate slows down, while Kohler (1983a) found an increase in the average F0 level for fast speech in his German material.

In summary, these few consistent findings based on read speech do not allow as many generalisations as were possible e.g. for pausing. But there are clearly some interactions between tempo and intonation, with a general tendency for more accents at a slower tempo and fewer accents at a faster tempo. The number of pitch accents has, of course, implications for the ryhthmical structure sketched in the next section.

3.3. Rhythm

Since it reflects how (in our case) speech is temporally organised, rhythm clearly has implications for tempo variation. Problems arise when, for some authors, rhythm is more or less synonymous with tempo and the timing of speech. Another problem lies in the fact that, for the author, rhythm seems, paradoxically, to be one of the most obvious and at the same time one of the most controversial topics in phonetics and phonology. This seems particularly true when it comes to concrete temporal correlates of rhythmic units.

The classification of the languages of the world in stress-timed and syllable-timed languages by Pike (1945) led to the hypothesis that basic rhythmic units in a given language have approximately the same duration. However, this isochrony hypothesis has consistently failed when the actual durations of feet (stretches between two stressed syllables) in the stress-timed languages, or syllables in the other language rhythm-type were measured. Neverthless, there is a tendency to shorten syllable durations in feet with many syllables, and vice versa, evidence which can be seen as a sort of rhythmic compensation.

Several researchers see a tendency to isochrony in speech production in stress-timed languages such as English (Lehiste, 1975) and German (1982). In the latter case the isochrony tendency is manifested not only by durational compression but also in phonemic reductions (see next section) and modifications of the number of syllables

and of the syllable complexity (see section 6.5) and the accent patterns (cf. last section). We therefore avoid the traditional complexities of the rhythm discussion and address these phenomena in the appropriate sections.

3.4. Segmental processes of connected speech

Fast speech is often characterised by processes on the sound segment level, such as deletion, assimilation and replacement of those sounds that are defined in the canonical or lexical structure of the words. These processes happen in connected speech and are normally treated as "post-lexical rules" by phonologists. The example in table 3.2. shows various segmental forms of the same sentence. Although the processes also occur in normal-tempo speech, the resulting segmental chain for fast speech differs from the one expected at a normal tempo. The corresponding speech signal shows a different stream of acoustic phonetic segments.

Table 3.2. The German example "Hast Du einen Moment Zeit?" in possible various forms starting with the abstract underlying form, the canonical realised form to more and more reduced forms (after Kohler, 1990).

full lexical (underlying) form	h a s t d u: aɪ n ə n m o m ɛ n t ts aɪ t
canonical form; assimilation voice; degemination	h a s **t** u: ʔ aɪ n ə n m o m ɛ n **t**s aɪ t
reduction of vowel quality, deletion of schwa; degemination	h a s t **ʊ** aɪ **n** m o m ɛ n ts aɪ t
reduction of vowel quality	h a s t ʊ **ə** n m o m ɛ n ts aɪ t
merging of vowels with reduction of vowel quality; assimilation of place of articulation with degemination	h a s t **ə m** o m ɛ n ts aɪ t
Deletion of schwa; assimilation of place of articulation	h a s **p** m o m ə n ts aɪ t

The canonical or lexical form of a word can be altered in different ways and at all levels of style and tempo. Deviations from the lexical form comprise basically assimilations of all types and complete deletions of segments. It is evident that in fast

speech more deviations from the lexical form occur, so that sometimes the term "fast speech rules" or "allegro rules" is used.

These allegro rules are sometimes used to characterise certain speaking styles which are denoted as casual, informal or sloppy, as well as for diachronic phonological processes (Dressler, 1975). Of course, these processes have been observed and described for languages other than German, such as for Hebrew (Bolozky, 1977) to name just one example.

Greisbach (1992) tested the processes mentioned e.g. by Kohler (1990) for German speech read at maximal speed. He indeed found all the described processes and added some new rules to the ones already described. It is noteworthy that the realisations of the five speakers differed considerably in terms of rule selection.

Function words, which frequently occur in their weak forms deserve special consideration. The weak forms often show a tendency to reduce the vowel to a schwa, cf. "can" as [kæn] and [kən], "for" with [fɔː] and [fə]. Although there is a permanent propensity to use weak forms, the degree of "weakness" and the frequency of weak form occurrence seems to increase in styles marked as fast, or as Kohler (1995: 220) notes:

> "Je schneller das Redetempo ist, desto leichter stellen sich schwache Formen ein und desto weitreichender sind auch die Veränderungs-prozesse."

3.5. Segment and syllable duration

A change in the articulation rate does, of course, imply a change in the temporal extension of speech stretches corresponding to linguistic elements such as syllables and sound segments. But as already pointed out with the change of pauses and the change of sound segmental processes, the change of the durations of sounds and syllables does not apply in a linear way.

It has been observed that sound segments reveal different behaviour in terms of compressing and expanding their durations according to their sound class. This phenomenon has been described as the *elasticity of sounds* (Gaitenby, 1965; Campbell & Isard, 1991). Roughly speaking, vowels expand and compress more than consonants, i.e. a slow articulation means primarily vowel lengthening, and fast

articulation means primarily vowel shortening. However, there are many factors responsible for segment durations.

Extra-linguistic and para-linguistic factors influencing speech tempo have been presented in the previous chapter. There also, language-relevant factors were listed which are not genuinely of a phonetic or phonological nature. The phonetic or phonological factors which can influence the duration of sounds and syllables are presented in this section. Table 3.3 gives an overview of various factors influencing segment duration which have been reported in the literature for different languages. It must be noted that the studies and languages mentioned represent only a small selection out of a rich pool of research literature on the topic of speech segment duration.

The durations of the acoustic correlates of sound segments show great variability, even among the realisations of the same phoneme. Durational variability has also been found between speakers as well as within speakers (Klatt, 1976).

On the sound segmental level, an inherent duration of segments has been observed in the way that e.g. close vowels show shorter durations than open vowels (cf. the literature presented in Lehiste, 1970). The phonological quantity (sometimes expressed as degree of tenseness) can have an immense influence on the actual duration of segments, e.g. underlyingly long and tense vowels in German show longer average durations than short and lax vowels.

On the syllabic level, the number of segments (especially the number of consonants) have an influence on the duration of all segments in that syllable. In "strict" with five consonants, the durations of all sounds are expected to be shorter than in "tick" with only two consonants.

But also the position of the consonant in the syllable can make a difference for the consonant duration. The duration of [f] in syllable coda position (e.g. in "Schiff", engl. "ship") is expected to be longer than in syllable onset position as in "Fisch" (engl. "fish").

For vowel duration the (phonological) voicing status of the following consonant can be decisive. If the post-vocalic context is a lenis plosive like in "bag", the vowel is probably longer than in "back" with a fortis plosive as post-vocalic context. Despite word-final devoicing in German, this also applies where the fortis-lenis opposition is maintained (e.g. "leiten" vs. "leiden").

On the word level, the number of syllables in a word makes an essential contribution to the duration, especially for vowels. The [ɪ] in "stick" is longer than in "sticky" and the vowel in "stickily" is shorter than either of the others (Lehiste, 1972).

Similarly, the position of the syllable in the word can determine the segment duration. Word-final lengthening was observed e.g. for Dutch (Nooteboom, 1972), Swedish (Lindblom & Rapp, 1973) and American English (Oller, 1973; Beckman & Edwards, 1990).

One of the main factors determining duration is the lexical stress as the prominence on the word level. On the prosodic phrase level, the lengthening of pitch accented syllables, in addition to their inherent lexical stress, is a further factor. This illustrates the hierarchical and cumulative effects of duration-influencing factors.

A further prosodic condition leads to considerable durational changes, notably the lengthening effects at edges of prosodic constituents. The best known effect is phrase-final lengthening. But a phrase-initial strengthening with consequences for duration has also been reported (Fougeron & Keating, 1997).

Table 3.3. Factors which influence segment duration with evidence for different languages from a selection of studies. Studies that also looked for tempo are indicated with an asterisk (*).

factor	**study**
inherent durations	American English: Lehiste (1970); Klatt (1975); German: Neweklowsky (1975); Antoniadis & Strube (1984); French: O'Shaughnessy (1981)
phonological quantity	Amer. English: House (1961); German: Antoniadis & Strube (1984); Braunschweiler (1997)
consonant cluster	American English: Klatt (1975); German: Kohler (1988)
position in syllable	American English: Klatt (1975); Greek: Botinis et al. (1999)*; German: Kohler (1988)
pre-, post-vocalic context	Amer. English: Peterson & Lehiste (1960), House (1961), Lisker (1974); German: Antoniadis & Strube (1984); French: O'Shaughnessy (1981)
polysyllabic shortening / monosyllabic lengthening	Dutch: Nooteboom (1972); Swedish: Carlson & Granström (1986), Lindblom & Rapp (1973); German: Kohler (1986); American English: Klatt (1975)

position in word / word-final lengthening	Dutch: Nooteboom (1972); Swedish: Lindblom & Rapp (1973) ; Amer. English: Oller (1973); Beckman & Edwards (1990)*; French: O'Shaughnessy (1981)
word-initial consonant lengthening	Dutch: Nooteboom (1972); Swedish: Lindblom & Rapp (1973); Amer. English: Lehiste (1960); Oller (1973)
lexical stress	American English: Klatt (1975); Botinis et al. (1999)*; German: Jessen et al. (1995)
pitch accentual lengthening	German: Kohler (1988); English: Turk & White (1999); Dutch: Eefting (1991)
phrase-final lengthening	Amer. English: Gaitenby (1965); Klatt (1975); Lehiste, Olive & Streeter (1976); Beckman & Edwards (1990)*, Turk (1999); Hebrew: Berkovits (1991)*; Dutch: Gussenhoven & Rietveld (1992); French: Fougeron & Keating (1997)
phrase-initial strengthening	French: Fougeron & Keating (1997)
foot shortening	Dutch: Nooteboom (1991); Amer. English: Beckman & Edwards (1990)*

There is still much to discover about the domains in which these factors operate. The factors listed in table 3.3 are considered to result in local changes of articulation rate. A linguistic factor affecting global tempo is utterance length. A number of studies have shown that the length of an utterance has an influence on the tempo of the utterance. Fónagy & Magdics (1960) for Hungarian and Malécot et al. (1972) as well as Bartkova (1991) for French give evidence for the tendency that speech rate increases with length of utterance. Of course a long utterance would normally show fewer prosodic phrase boundaries than two or more short utterances, and that means fewer phrase-final lengthened syllables. Nevertheless, Haselager et al. (1991) found in their study with children, where they disregarded pauses, the vowel and the consonant(s) before the phrase break, that longer utterances are articulated at a faster rate than shorter utterances. Thus, it can be hypothesised that we have a shortening effect such as the shortening due to an increased number of segments in a syllable, or due to an increased number of syllables in a stress group. This shortening effect has been demonstrated e.g. by Lehiste (1972). In her material, not only utterance duration (in msec) and utterance length increase with the number of syllables and segments, but also the syllabic rate, with one word sentences the slowest and six word sentences the fastest utterance. Gaitenby (1965) makes the following observation in her sentence material: the longer the utterance in terms of number of segments, the shorter the absolute duration of any given segment, until an approximate minumum duration was

reached beyond which segments could not be compressed further. This shortening effect has also been observed in natural recordings. In a longitudinal study of adult-child interaction, Van de Weijer (1997) offers evidence that articulation rate continously increases from utterance span of one syllable up to seven or more syllables, for both child-directed and for adult-directed speech.

All factors can interact with each other, and these interactions must be considered as well in a duration model. There are studies investigating some of the interactions of the mentioned factors with rate. Berkovits (1991) found a phrase-final lengthening effect which only operated with fast speech, not for slow speech in her Hebrew data. Thus, a duration model which attempts to model the effects of articulation rate as one of the main factors of durational variability must take into account the main potential interactions with the other factors.

3.6. Articulatory organisation

When humans articulate faster, different mechanisms of articulatory organisation may come into play. This can be achieved e,.g. by *shortening the duration* of the articulatory gesture, as discussed e.g. by Kröger (1996) and Gay (1981).

The best known mechanism applicable to fast speech is the *target undershoot*, i.e. a reduction in the magnitude of articulation. The theory of target undershoot (Lindblom, 1963) says that in a shortened sound segment the articulatory, and consequently the acoustic target has not been fully reached before the particular articulator starts the next gesture. The resulting spectral reduction of time-reduced vowels is expressed as a tendency for centralisation, i.e. that they are more central in the vowel space. Regarding different forms of shortening (induced by different tempo and different degree of stress, respectively), he found in his study support for the hypothesis that

> „it is immaterial whether a given length of the vowel is produced chiefly by the tempo or the degree of stress. Duration seems to be the main determinant of the reduction.“

Contrary to Lindblom (1963), other reserachers were not able to find evidence for target undershoot in fast speech conditions (Engstrand, 1988; van Son & Pols, 1989; Nooteboom 1991).

A third mechanism for faster articulatory movement is the *increase of velocity*. The data investigated by Kuehn & Moll (1976) and Gay (1981) show that the velocity of articulators can increase in fast speech.

A further mechansim to speed up articulation is to *increase the gestural overlap*. Adjacent sound segments that use different articulators can also overlap in production. For example, [t] does not require the use of the lips, so lip rounding in an adjacent segment (like [u:]) can begin during the [t], confer the first [t] in "tourist" with the one in "tick". As an example study, Engstrand (1988) varied stress and tempo in vowel-consonant-vowel sequences with Swedish speakers. He found that vowel- and consonant-related gestures were coproduced to a greater extent at fast tempo compared to slow tempo.

Also the *degree of coarticulation* can vary as a function of speech rate. In coarticulatory assimilation, neighbouring sound segments that require the same articulator use a single articulatory gesture for both sounds. E.g. in the phrase "Er hat ja gelogen.", [t] and [j] both require particular articulations of the tongue tip and blade. In this case, [t] is often produced with the palatal gesture of the [j], resulting in a [ç]-like release of the plosive.

Although the mechanisms are mentioned separately they do not necessarily apply separately. In an American English study investigating different consonant clusters under different speech rate conditions eliciting electropalatographic data, Byrd & Cheng Tan (1996) report individual consonant shortening in duration and a relatively increase in the overlap of the articulations.

To summarise, the mechanisms mentioned above - as well as those mechanisms on the other structural levels – are non-linear in nature. This phenomenon has also been recognised as a general principle of articulatory tempo variation by Gay (1977):

> "The reduction in duration of all segments coupled with the relative constancy of acoustic (vowel) targets, suggests that this adjustment [of articulatory movement] involves primarily a horizontal compression. This [horizontal] compression [...] is a non-linear one, and one that causes both a decrease of duration and an increase in coarticulation."

Summary and discussion of chapter 3

A change of tempo results in changes at *many* levels of phonetic and phonological characterisation, and not only at one level as one might think if just pausing or articulatory velocity of a specific articulator is the subject of study. These levels can be more or less closely linked to each other, e.g. phrase break and segment duration vs. phrase break and pitch accent. Thus, tempo variation is very complex seen from the articulatory point of view.

It seems a general principle at all levels that changes in acoustic duration resulting from changes of speech tempo are *non-linear*. This fact makes the modelling of speech tempo much more complicated than a simple model based on a combination of linear changes.

If speech tempo is to be modelled, then it is necessary to develop the model at the different levels presented above. If speech tempo is to be modelled in a non-linear way, as observed in natural speech, then knowledge must be acquired about the non-linearity, and the magnitude of change at each level must go into the model. And finally, if speech tempo is to be modelled non-linearly for a speech synthesis application, then the general model based on speech production must be adapted for a given artificial speech generation architecture and the performance of the implemented model should be tested for speech perception, i.e. with actual listeners.