Chapter 2

Sources of Tempo Variation

Introduction

Why and where do speakers change tempo? There are an infinite number of reasons why, and situations where we vary our speaking rate or can observe, more or less consciously, different tempos in other speakers.

In order to structure this discussion, the distinctions *extralinguistic - paralinguistic - linguistic* are chosen. Although there are no sharp boundaries between these three terms they help to illuminate different levels of the problem. Some aspects can be attributed to the individuality of a given speaker (*extralinguistic*). Other aspects can be explained solely by the situation and/or the inner state of the person who is speaking (*paralinguistic*). Third, other aspects can be unambiguously attached to how spoken language is performed in interaction and in order to convey verbal information (*linguistic*). The current chapter seeks to show the diversity of factors with observations from production as well as from the perceptual perspective. The subsequent chapter 3 is devoted to the encoding and the execution of the *phonetic plan*, i.e. how the phonological encoding is structured and how the resulting phonetic plan is realised in articulatory actions leading to actual speech.

Even though the present chapter cannot claim to represent a *complete* list of sources of tempo variation, it gives an idea of the range of sources of variability and it shows that attempts to explain variance in the linguistic-phonetic expression are poor if paralinguistic and extra-linguistic factors are not included. The "neutral" situation in recording laboratories simulating communication is not the same as communication in the real world. Lab-speech experiments can help to explain how speech communication may work, but only to a limited extent with respect to real speech.

2.1. Preliminary explanations

It is necessary to define some central terms and concepts at the very beginning to avoid a terminological confusion. These terms occur throughout the thesis and will be explained in detail with each chapter.

There are various terms used to denote the tempo of speaking such as *speech rate*, *rate-of-speech (ROS)*, *rate of speech production*, *speed of talking*, *talking rate*, *reading rate* (for read speech), *speaking tempo* or simply *tempo*. These terms are used here as synonyms and most of the time in this thesis the expression *tempo* is used.

The pause plays a central role when dealing with tempo and it frequently makes a big difference if pauses are taken into consideration or not. The usual distinction is that tempo can either be defined as articulation rate or as speaking rate. *Articulation rate* as a net rate refers to phases of articulation ***ex****cluding pauses*. *Speaking rate* as a gross rate refers to the entire speaking phase ***in****cluding pauses*.

Many expressions have been "invented" for *articulation phases* demarcated by two pauses: "chunk" (Fougeron & Jun, 1998), "run of pause-free speech" (Miller, Grosjean & Lomanto, 1984), "interpause stretch" (Dankovičová, 1997); "run" (Crystal & House, 1990), "articulatory run" (Tsao & Weismer, 1997), "interpausal speech run" (Koopmans-van Beinum & van Donzel, 1996), "phrase" (Fant, Kruckenberg, & Nord, 1992), "utterance" (Butcher, 1981), "T-phrase" (Eefting, 1991) and "speech chain" (Bartkova, 1991). The term used here is *inter-pause stretch*, because it seems the most informative.

Usually we want to categorise the tempo, i.e. whether a speech sample is considered fast, or slow, or slower than normal, or whatever the intended relational purpose might be. It must be emphasised that this categorisation strongly depends on the phonetic perspectives of speech production, speech acoustics, or speech perception. Although a speaker intends to speak "fast", the resulting speech will not necessarily be categorised as "fast" on the basis of a physical measurement (e.g. in syllables per second). Furthermore, this stretch of speech will not necessarily lead to the auditory impression of "fast" for listeners. Thus, tempo categories are only comparable and interchangeable under certain conditions. And it is with the same caution, that production studies have to be compared with perception studies in the review chapters.

The last point to be made in advance is that we can distinguish between the *subjective* tempo in production and perception, and a measured *objective* tempo. It is without doubt useful to have a standardised metric to quantify tempo. But bear in

mind that there *is no* such standard, and that there *cannot* be such a standard as will be explained in chapter 4. The reasons are, among others, differences in the definition of the linguistic units, usage of pause, material and structure of language. Although there is no exact standard, similarly used measurements express similar things *more or less*. With this sensitivity in mind, the reader is referred to table 2.1. It contains a list of various studies dealing with different material in several languages where articulation and speaking rates were measured in syllables per second as the most popular tempo metric.

Table 2.1: Survey of studies investigating speaking rates in different languages and accents. Speaking rate (sr = including pauses) and articulation rate (ar = excluding pauses) is indicated in syllables per second for two different speaking modes: reading and spontaneous. For articulation rate the percentage of pause time of the whole speaking time is given. Numbers with an asterisk (*) are re-calculated as syll/sec for the number of speakers for the material on the basis of the data in the literature. Numbers with a double asterisk (**) indicate re-calculations either from the pause quotients or from the relationship between speaking rates and articulation rates in the original studies. For studies with various rates, only the data for the "normal" or "medium" rate are used here.

		subjects		**read**			**spontaneous**		
study	**text type**	**no.**	**language**	**sr**	**ar**	**pau**	**sr**	**ar**	**pau**
Dauer (1983)	prose	1	Engl. (UK)	5.9					
		1	English (US)	5.0					
Iivonen et al. (1995)	news	8	Engl. (UK)	5.3	5.4				
		9	English (US)	5.2	5.4				
Tsao & Weismer (1997)	prose	100	English (US)	4.39					
Hewlett & Rendall (1998)	neutral text & conversation	12	Engl.(Orkney)	4.50	5.49	18%	4.53	6.02	29%
		12	Engl. (Edinb.)	4.55	5.43	16%	4.34	5.52	24%
Tauroza & Allison (1990)	news / radio announcements		English (UK)	4.16					
	conversat.		English (UK)				4.39		
	interviews		English (UK)				4.18		
	lectures		English (UK)				3.24		
Grosjean & Deschamps (1975)	radio interviews	30	English					5.17	

Grosjean & Deschamps (1975)	radio interviews	30	French					5.29	
Fletcher (1987)	transribed interviews	6	French	4.49	5.53	19%			
Malécot et al. (1972)	conversations	60	French (Paris)				5.73		
Fougeron & Jun (1998)	prose	3x3	French (Paris)	4.32	5.65				
Slembek (1993)	news	2	French (FR)	5.32		11%			
	news	2	French (CH)	5.04		7%			
Slembek (1993)	news	2	German (CH)	4.24		15%			
	news	2	German (DE)	4.84		17%			
Meinhold (1967)	prose	14	German	5.40*		29%*			
	news	17	German	5.73*		18%*			
	poetry	8	German	3.63*		30%*			
Greisbach (1992)	texts	8	German	4.81*					
Iivonen et al. (1995)	news	5	German	5.80	5.90				
Künzel (1997)	news magazines	10	German	5.12	6.04	15%			
	monologues	10	German				4.28	5.89	29%
	monologues	10	German				4.18	5.83	29%
Trouvain (1999)	transcr. news	3	German	4.72	5.30				
	prose	3	German	4.67	5.52				
Wiese (1983)	cartoon retelling		German					4.57	
Strangert (1993)	transcr. news	1	Swedish	5.77	7.98	28%	4.17	7.80	46%
	news	1	Swedish	5.83	6.4	9%			
Koopmans-V.B. & Van Donzel (1996)	retold story	8	Dutch				5.85	3.79	35%
Iivonen et al. (1995)	news	18	Finnish	6.3	6.5				
Dauer (1983)	prose	3	Spanish	7.10					
Dauer (1983)	prose	3	Greek	7.47					
Dauer (1983)	prose	2	Italian	7.30					

2.2. Extra-linguistic sources of tempo variation

Habitual speech rate

Individual speakers can differ substantially in their typical speech rates. This can easily be shown by looking at a database containing the same text read by many different speakers. In the German "Kiel Corpus of Read Speech" (IPDS, 1994) 16 speakers read the IPA standard text "Nordwind und Sonne". The descriptive statistics of the speaking rate in table 2.2 show a considerable variation across the readers (half a syllable per second standard deviation) with the slowest reader more than two syllables per second slower than the fastest reader. Similar results were reported for read German speech (news magazine) by Künzel (1997), and have also been observed for other languages such as English (Goldman Eisler, 1968).

Tsao & Weismer (1997) investigated the habitual and the maximium articulation rates in extremely slow and extremely fast speakers (+/- 1 standard deviation from mean in a reading task). The results reveal that the slow readers at their maximum articulation rate could just articulate as fast as the fast readers at their habitual articulation rate. Since the speeding up magnitudes for both groups behave similarly, there are indications that the maximum rate for an individual can be predicted from his/her habitual rate.

Table 2.2. Mean speaking rate and mean articulation rate, with its standard deviations (sd), maximal (max) and minimal (min) values measured in underlying syllables per second of two German data collections: the "Nordwind und Sonne" recordings in the Kiel Corpus (excluding the pause between the two paragraphs) and readings of news magazine articles (Künzel, 1997).

	no.	**speaking rate**		**articulation rate**	
corpus	**speakers**	**mean (sd)**	**min-max**	**mean (sd)**	**min-max**
Kiel	16	4.27 (.51)	3.05 - 5.18	*to be added*	*to be adde*
Künzel	10	5.12 (.42)	4.52 - 5.82	6.04 (.50)	5.31 - 6.90

Age

Haselager et al. (1991) investigated articulation rate skills of Dutch-speaking boys and girls in four age groups between 5 and 11 years. They used a diadochokinetic task (repeating the same syllable as fast as possible) as well as spontaneous speech elicited in interviews. For both speaking modes the syllabic rate varied with the age group: the younger the group the slower the articulation rate. The same effect has been observed by Walker et al. (1992) for English speaking Canadian preschool children. Their results from a spontaneous speech and a speech imitation task show significant differences in the articulation rate between children at age 3 and at age 5. The age effect is also reported in a British study by Whiteside & Hodgson (2000). In a picture-naming experiment with children aged 6, 8, 10 years and an adult control group they found significant differences relating to articulation rate as a function of children's age. There is evidence that the increase in articulation rate during maturation proceeds in a non-linear way as pointed out by Hall, Amir & Yairi (1999).

An age effect seems to apply not only to the developmental phase of speakers. Malécot et al. (1972) report in their study on French spontaneous speech that older adults speak slower than younger adults: syllabic rate drops progressively by about half a syllable per second overall from 5.95 syll/sec to 5.52 syll/sec, for young (starting at 20) to older speakers (up to 69 years).

These findings are backed by an American English study by Sommers, Humes & Pisoni (1994) who investigated the effects of increased speaking rate and greater speaking-rate variability on spoken-word recognition in older and younger listeners. For younger subjects, neither increased speaking rate nor greater rate variability produced significant changes in perceptual identification scores. Older listeners, in contrast, exhibited significantly poorer identification scores for fast, compared to medium or slow speaking rates. In addition, trial-to-trial variations in speaking rate produced a significant decrease in identification scores for elderly subjects listening to fast-rate item.

Gender

In Whiteside & Hodgson (2000) a significant difference for gender was found (except for the six years old): females articulate slower than males. This confirms the findings in Whiteside (1996) for read sentence material. American variants of English also exhibit gender differences for tempo, as has been shown by Byrd (1992) for the TIMIT database. Considering only vowel duration, women exhibit longer values than men (Simpson, 1998, for German; Simpson, 2001, for American English). However for French, Malécot et al. (1972) found *no* significant differences between the sexes in terms of syllable rate, but there *were* differences in terms of utterance length. The findings of an English study (Deese, 1984) contradict the aforementioned studies. Here, women spoke faster (5.82 syll/sec) than men (5.48 syll/sec) but this difference was not statistically validated. Thus, whether gender influences tempo and articulation rate remains an open question.

Speech and hearing impairments

Apparently, most speech and hearing impairments also have an effect on speech tempo. In the area of motor speech disorders, tempo may be slowed down or shows great variations as has been described for the continuum of developmental dysarthria to developmental verbal dyspraxia (MorganBarry, 1995). These articulation disorders with neurological origin also show other prosodic symptoms such as erratic pausing, arhythmic structures, unfinished intonation units and disfluencies. The marked feature of various forms of stuttering (or stammering) is the abnormal number of disfluencies such as arhythmic pausing, blocking of articulatory air-flow, prolongations of sounds, restarts and repetitions of sounds, syllables and words. These disfluencies makes the overall tempo rather slow. Disfluencies such as repetitions, repairs, and filled pauses usually occur in spontaneous speech of non-stutterers. The moderate number and the type of disfluencies seem to count as criteria to classify them as "fluent". It is interesting to see here, that in fluent phases, persistent stuttering pre-school children show no significant difference with respect to articulation rate to a non-stuttering control group (Hall, Amir & Yairi, 1999). With regard to stuttering it is interesting to note that parental high speaking rate results in a degraded fluency in the children. On the other hand parental slowing often leads to an improved fluency, maybe because parents also

show changes in behaviour in addition to slowing down speech rate. This can reflect more empathy with their children, as hypothesised by Guitar & Marchinkoski (2001).

Special forms of speech and hearing impairments are those caused by the use of drugs. As an example, alcoholic intoxication has an effect on speech rate as reported by Künzel et al. (1992) in a reading task: speakers under alcohol articulated more slowly and made more and longer pauses, including a greater number of hesitation pauses.

In a study investigating the intelligibility of sentences spoken in a conversational style and at a fast rate, and those spoken in a clear style and at a slow rate, Uchanski et al. (1996) found that listeners with a hearing loss show significantly better results for the clear speech condition (87% vs. 72% for conversational). Listeners with normal hearing show only a slight increase (98% vs. 92%) for the clear condition. This shows that speaking slower and more clearly to hard-of-hearing persons makes understanding easier for them.

Auditory conditions

In the same study by Uchanski et al. (1996) the same conditions (conversational/fast vs. clear/slow) were also tested with normal hearing listeners under noise conditions simulating a hearing loss. Under these adverse auditory condition the intelligibility effect of clear and slow speech is even more evident than for the hearing loss condition: 60% for clear vs. 44% for conversational style. These results give rise to speculations that under anything less than ideal listening situation clear and slow speech contribute significantly more to intelligibility than "natural" conversational and fast speech. That means that in unfavorable listening conditions – and here synthetic speech must be included – listeners would prefer slower speech.

Cultural and geographical background

Similar to dialectal and sociolectal phonetic differences one could imagine that differences in tempo occur between speakers of the same language but with a different cultural or geographical background. But Slembek (1993) found no tempo differences between broadcast news readings from stations of the different nations. French speaking news readers in France do not differ in syllabic rate from those in Switzerland, and

Swiss German radio news readers speak as fast as their colleagues in Germany. In a similar study Iivonen et al. (1995) found no tempo differences between American and British English radio news. However, Byrd (1992) reported in her analysis of the TIMIT database some statistically significant differences in tempo and pausing between speakers from different dialect regions in the United States.

Hewlett & Rendall (1998) investigated the question of whether lifestyle, in the form of urban vs. rural living, influences speech rate. Comparing Scottish English speakers from Edinburgh with those from the Hebrides, they rejected the claim that is sometimes made that city residents speak faster than those living in the countryside.

Language proficiency

Different languages may possibly differ in terms of rate of speech production units as can be seen in the studies for different languages in table 2.1. But there are certainly differences in terms of how speech rate is perceived across languages. Speech of the native language/s or those which are mastered with a higher level of proficiency is felt to be less fast than the speech of those languages with a lower level or no proficiency. Abercrombie (1967: 96) puts it as follows:

> "Everyone who starts learning a foreign language, incidentally, has the impression that its native speakers use an exceptionally rapid tempo."

In a study comparing spontaneous monologues of American English and Japanese speaking students Osser & Peng (1964) found no significant differences in terms of the phoneme production rate of the two groups. They explain the impression that an unknown language sounds faster than normal, i.e. than one's own language, with phonological differences such as different patterns of syllabic complexity: Japanese people tend to perceive unknown consonant clusters of English as syllables and therefore the number of perceived English syllables increases which results in a higher perceived syllabic rate. In contrast, English speaking people tend to interpret the many vowels and the many syllables but relatively few phonemes in Japanese as a higher syllabic rate compared to English.

Speaking a foreign language is usually linked with a higher cognitive activity. The problems include incompletely developed syntactic and morphological knowl-

edge, slower lexical access, and articulatory difficulties in less well established segmental and prosodic patterns. The process of planning and executing speech is slowed down and this is normally mirrored by reduced fluency in the non-native talker. Evidence for this claim is given in e.g. Pürschel (1975) who examined German students of English. The students were first asked to read an English text and afterwards the same text in a German translation. There were more pauses in the reading of the foreign than in the native language text leading to a slowed down tempo. Wiese (1983) also investigated the temporal behaviour of German learners of English and native speakers of English with a cartoon retelling task. He found significant differences in terms of mean pause duration (slowing down the overall speaking rate) as well as in terms of mean articulation rate between the two groups. It seems that speaking in a foreign language means articulating more slowly and making more and also longer pauses than usual, i.e. than in the native language.

But *comprehension* of a foreign language is also affected by speech rate. As an example, Griffiths (1990) tested the comprehension of Japanese teachers listening to English texts delivered at three different rates. The moderately fast readings resulted in a significantly lower comprehension score than the slower readings. Anybody who has ever tried to learn a new language can probably confirm these findings. Therefore it is not surprising that learners sometimes explicitly ask for slowed down speech. This wish for special listening conditions is fulfilled e.g. by the Deutsche Welle, the German broadcast station abroad, that offers an additional version of slowed down broadcast news for foreigners (Deutsche Welle URL) where the news is spoken with a speaking rate of about 3 syllables per second compared to the usual speaking rate span of between 4.5 and 5.5 syllables per second for German broadcast news.

Apart from the fact that "normal" native speakers' tempo appears as "fast" for language learners (L2 judge L1 speech tempo), and that language learners produce the foreign language at a rather "slow" rate compared to their native language or to the tempo of native speakers (comparing L1 and L2 speech tempo), the tempo of language learners has an effect on the proficiency judgements by native-speaking listeners (L1 judge L2 speech tempo), in addition to segmental and prosodic errors. Munro & Derwing (2001) asked English native speakers to judge L2 speech (L1: Mandarin) on accentedness and comprehensibility. The articulation rate of the read sentences was manipulated with a speech compression-expansion editor by 10 % so that each sentence was presented in a slightly slower version, the actual version, and a slightly faster version. As expected the native-speaking listeners evaluate slightly faster versions as less accented and more comprehensible than the natural or even the slowed down tempo of foreigners' talk. However, talking too fast resulted in a downgrading.

Moreover, the study shows that tempo makes a small but significant contribution to both accent and comprehensibility ratings independent of segmental errors.

2.3. Paralinguistic sources of tempo variation

Emotions

Emotions can have a strong effect on speech tempo. Expressive speech is marked by global prosodic parameters such as F0 variation, voice quality and speech tempo. Several studies show evidence of general tendencies for some given emotional categories such as anger, joy or sadness (for an overview see van Bezooyen, 1984; Murray & Arnott, 1993; Banse & Scherer, 1996; Burkhardt, 2001). Although language, exploration methods, purpose, and speech material differ in these studies, the reported patterns look alike. Anger, rage, fear, but also happiness is generally marked by an increased tempo whereas boredom, sadness, sorrow, grief, and disgust is characterised by a slowed down tempo.

An alternative way of describing emotions is along three dimensions rather than with labels for "full-blown" emotions. The three dichotomies are dominant-subordinated (dominance), positive-negative (valence), and active-passive (activity). Speech tempo seems to correlate strongly with the activity dimension (Schröder, in prep.; Kehrein, 2002), i.e. the more active the speaker the faster s/he speaks, and the more passive the speaker the slower s/he speaks (Scherer, 1974; Apple, Streeter & Krauss, 1979).

Stress

Similar to emotional stress, cognitive stress can also result in high arousal. Lively et al. (1993) examined the effects of cognitive workload on speech production. Workload was manipulated by having speakers perform a compensatory visual tracking task while speaking short carrier sentences. In the workload condition, speakers produced utterances with increased amplitude and amplitude variability, decreased spectral tilt, increased F0 variability, and increased speaking rate.

Barber et al. (1996) also showed that a high cognitive workload leads to a faster tempo. In the two reported experiments (time-stress and dual-task performance) most of the English subjects doubled their speech rate. However, Dankovičová & Nolan (1999) were not able to show consistent speeding up when cognitive stress was present.

Competence and benevolence

Categories such as competence and benevolence have been related solely to perceptual impressions. Smith et al. (1975) synthesised utterances and manipulated them with respect to tempo. Then, subjects were asked to evaluate the stimuli on a list with attributes which could be summarised under the headings *competence* on the one hand and *benevolence* (cooperativeness, friendliness, politeness) on the other. The results show that the highest benevolence score correlates with the "normal" speaking rate, with a linear decrease to both sides, i.e. the more the speed increases or decreases the more the benevolence scores decrease. At the same time, the results show a tendency for faster rates to be given higher competence scores. In a similar study Apple, Streeter & Krauss (1979) summarise their results:

> "... slow-talking men are judged to be less truthful, fluent, emphatic, serious, and persuasive, and more passive, although they are also seen as more potent."

Ofuka et al. (2000) found in their comparisons of polite and casual Japanese utterances that speech tempo was used consistently by all six speakers: comparing polite with casual speech it appeared that polite speech was in general slower than the casual form. Apart from tempo and the F0 movement on the final vowel, the duration of the final vowel affected the politeness rating of Japanese judges. These ratings also revealed that the function of politeness related to speech rate was that of an inverted U-shape, i.e. the fastest and slowest versions get the lowest scores.

Communication partner

Everybody changes their speech tempo (among other prosodic parameters) more or less consciously when talking to people from whom one might assume less estab-

lished information processing abilities, including infants, non-native speakers, elderly people, or persons with hearing difficulties. The study of Van de Weijer (1997) examining the speech of the mother, the father and the baby-sitter of a Dutch child between age 6 and 9 months confirms the general findings of infant-directed speech. So-called "motherese" compared to adult-directed speech features a higher F0 average and a larger F0 range along with a slower articulation rate as well as shorter utterance length (in syllables) between pauses. In an investigation of speech addressed to elderly people (American English) Kemper (1994) comes to similar results: "elderspeak" is marked by a slower articulation rate, shorter utterances, and more pauses.

Frequently speakers unconsciously adapt their speaking rate to their dialogue partner's or to the assumed speech rate ability of their partner. The accommodation theory (e.g. Street & Giles, 1982) distinguishes full and partial convergence, maintenance, and divergence, whereby speech convergence represents a move towards social integration. The different categories are examined with various non-content speech behaviour parameters such as response latency, utterance duration and speech rate of two interactants. While everyday interlocution normally reaches a low level of awareness, the speech behaviour "divergence" is brought into consciousness and perceived negatively (Street, 1982).

Adapting speech rate to the communication partner has also been observed for children as early as 3 years old. Guitar & Marchinkoski (2001) observed in their study with mother-child dyads that in five of six cases the children significantly slowed down their tempo when their mothers spoke slower (on average by 51 %). Based on the positive correlations they discovered, the authors hypothesise that children also speed up when their parents speak faster.

2.4. Language-relevant sources of variation

Speech planning

In her summary of investigations about cognitive activities during spontaneous speech Goldman-Eisler (1968: 31) claims that spontaneous and read speech do *not* differ basically in articulation rate:

> "Variations in the overall speed of talking were found to be variations in the amount of pausing. What is experienced as increase of speed in talking proved to be variation in amount of pausing. The rate of articulation based on vocal activity exclusively, on the other hand, was shown to be relatively invariant."

According to her the only temporal difference lies in pausing, namely that spontaneous speech compared to read speech is characterised by *more* pauses, *longer* pauses and the presence of *filled* pauses. This different pausing strategy can be expressed as an increase of pause time as the portion of the total speaking time, which also leads to a slower speaking rate in syllables per second. The figures in table 2.1 (p. 7-8) confirm the different pause time ratios for various languages and various studies: between 5% and 20 % for read speech, and 30% and 46 % for spontaneous speech.

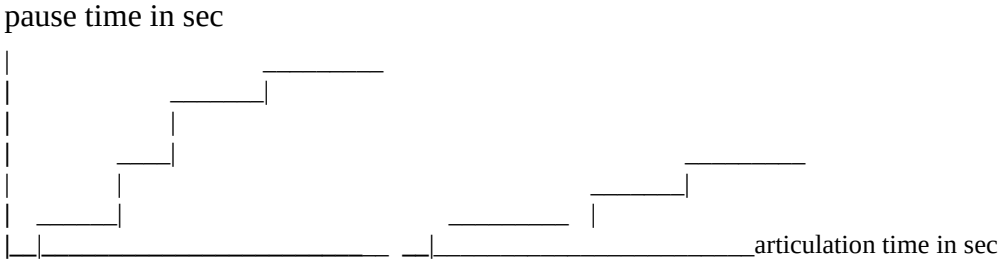
Read speech can be seen as a speech mode where the ideas to be expressed are completely prepared and formulated before speech production starts. In contrast, in many forms of spontaneous speech, the formulation process takes place "on-line" resulting in more pauses leading to a slower speaking rate. But the pauses seem to be unequally distributed over the utterance.

Levelt (1989:126) summarises the relevant studies:

> "There is some evidence that in longer monologues speakers slowly alternate between phases in which they spend much attention on information retrieval and inference (i.e. macroplanning) and phases in which they concentrate on finalizing messages for expression (i.e., on microplanning)."

The result is alternation of fluent phases (more articulation than pausing) and hesitant phases (more pausing than articulation) which reflects cognitive activity on the level of articulatory execution. The idealised scheme in figure 2.1 serves to illustrate these patterns of fluency and dysfluency in spontaneous and read speech styles.

Figure 2.1. Time course of spontaneous (left) and read speech (right) in articulating phases (x-axis) and pausing phases (y-axis). The degree of flatness mirrors the degree of fluency.



Let us ignore the pauses for now and consider only the articulation phases. Related to the quotation of Goldman-Eisler above, there seems no substantial differences in the *global* articulation rate between speakers or speaking styles. An articulation rate (e.g. in syllables per second) averaged over *all* articulation phases in the spontaneous mode of a speaker would not much differ from the average rate of *all* articulation phases in a read mode. However, the average rate does not tell anything about the variance *within* an articulation phase. In an investigation of interview responses by French speakers Miller et al. (1984) "discovered" this dynamic feature in the rate of articulation. They talk about "macro variables" which account for global tempo variation, in contrast to "micro variables" responsible for this local within-phrase tempo variation. As examples for these micro variables they suggest lexical access difficulties, syntactic construction delays and semantic planning problems.

All these "micro variables" point to some temporal delays during articulation typical for unscripted and more or less unplanned spontaneous speech, whereas read speech would lack these delays. In their Japanese study Hirose & Kawanami (2002) used dialogues simulated by actors which were also recorded as read isolated sentences. Although the above mentioned micro variables are not operating here, the dialogue speech samples show more dynamics than their read counterparts. These dynamics are expressed by the acceleration scheme between the two modes: compared to the readings, prosodic phrases in the dialogue turns were faster in the middle and slower at the end.

Types of spoken and written texts

Fónagy & Magdics (1960) measured tempo differences in text styles in their Hungarian study where sports news showed a higher tempo than read poetry. This is in line with Meinhold (1967) who found considerable tempo differences between readings of different text types, e.g. prose vs. poetry (see also table 2.1). Although the rate values for readings of prose and those of the news are close together, the results show only half as many pauses for the news (not in table 2.1) than for the prose text type whereas the mean pause duration is similar for both text modes. The data in table 2.1 give rise to the assumption that news reading has a faster speaking rate than other text types.

Abe (1997) investigated some prosodic characteristics of readings of different Japanese text types, such as a novel, advertisement phrases and paragraphs from an encyclopaedia. He found a much higher effect of pause and sentence boundary on the vowel duration of the preceding syllable for the novel compared to the other styles. The novel also showed the slowest speech rate, especially when pauses are included, whereas the encyclopaedia style is faster followed by the advertisement style which was the fastest.

In a study investigating recorded samples of ten different text types (such as prose, children's story, recipes, technical literature, dictionary) in three languages (Dutch, English, French) Fackrell et al. (2000) found that, in general, news readings are articulated faster than average and that dictionary entries, weather reports, and children's stories are read slower than normal for the three languages. However, the same tendency does not hold for each text type and language, e.g. advertising in English was faster than the average rate in contrast to a slower rate for the same text type in Dutch and French.

Similar to speaking styles based on texts, the styles of unscripted speech show variations of tempo. Kowal, Wiese & O'Connell (1983) performed a survey of various studies investigating spontaneous speech types such as descriptions of cartoons, pictures or films as well as speech in interviews (broadcast, television, medical patient). After a thorough recasting of all data available in five languages (German, English, French, Spanish, Finnish) they compared the monologic category of "storytelling" with the dialogic category of "taking part in interviews". It appeared that for both categories the articulation rate is comparable (5.17 syll/sec for storytelling vs. 5.26 syll/sec for interviews). However, in storytelling, pauses are made more often and are also longer, so that the percentage of pause time to total speaking time is greater (33%

vs. 17%) and the speaking rate is slower relative to interview speech (3.43 syll/sec vs. 4.31 syll/sec).

Tauroza & Allison (1990) showed in their British English data different speaking rates for various speech categories: turns in conversations are fastest (4.39 syll/sec), followed by radio announcements (4.16 syll/sec) and interview speech (4.18 syll/sec), and lectures being the slowest speech type investigated (3.24 syll/sec).

Dialogue management

A common phenomenon in a dialogue is that information is repeated, e.g. after a misunderstanding or to make something more explicit. One prosodic means is a slower articulation of the same word sequence and/or insertion of pauses, which also leads to a slowing down. This communication strategy is visible in an extreme form in man-machine communication: in German Wizard-of-Oz experiments the (human) operator of a dialogue system pretended to fail to understand (Fischer, 1999). Subjects showed a great repertoire of variation during the repetitions, depending mainly on the degree of cooperativeness. This variation in attitude is accompanied by various prosodic changes other than direct slowing down leading to a decreased rate, for example emphatic accentuation, more accents and hyper-articulation. These findings were confirmed in a similar experiment with synthetic speech for Viennese German (Pirker & Loderer, 1999).

In a dialogue, speakers are continuously sending and receiving signals on the status of the information exchanged. Confirmations and disconfirmations in the kind of an echoing response are usually marked by various prosodic means such as pause, duration, intonation contour and pitch range. As an example, Krahmer et al. (2002) found in their Dutch study that disconfirmative utterances were spoken more slowly than their positive counterparts, and that this feature was reliably used by listeners to classify those utterances as a negative response without context.

The study of Wells & Peppé (1996) can serve as an example of how tempo is used for turn organisation in dialogues. They found that in the Ulster variety as well as in the Tyneside variety of English a dialogue turn is delimitated by a markedly slowing down over the last two rhythmic feet (with a foot as a stretch of speech beginning with a stressed syllable). This turn-final lengthening is accompanyied by changes of other prosodic and non-prosodic phonetic properties such as loudness, voice quality, vowel quality and pitch contour.

Koiso, Shimojima & Katagiri (1998) claim for conversational Japanese that changes in tempo by the dialogue partners have a potential for cueing the structure of information collaboratively. In their data, openings of new information were marked by decelerations and the absence of information openings by accelerations.

Information management

Spoken language always features a certain degree of redundancy: it is not always necessary to understand all words to get the message. Greenberg (1999) speculates that

> "it is likely that frequently occurring words tend to be spoken faster and in more reduced fashion because of their inherent predictability."

This is well-known for the so-called function words (e.g. determiners, pronouns, auxiliary verbs, prepositions, interrogatives, conjunctions, degree adverbs) but also applies to frequently used lexical items. Normally, high-frequency words such as numbers are produced faster. In contrast to this, telephone numbers with their very low predictability are often spoken in a very slow way. If you miss a number or the correct order of the numbers you miss the whole message. Telephone numbers show *hardly any* or *no* redundancy. Thus, telephone numbers are optimally nested in a characteristic prosody which also features a slow speaking rate in terms of syllables per second as was shown by Baumann & Trouvain (2001) for German.

Uhmann (1989) found in her German data of everyday conversations that side comments such as parentheses and afterthoughts are marked by fewer pitch accents and a faster articulation whereas emphasised discourse segments are marked by more pitch accents and a slower articulation. This is in agreement with the analysis of Barden (1991) who showed for German dialogue speech that portions containing less central and less important information are spoken at a tempo faster than average, and inversely, that more central and more important portions are spoken at a slower speed than normal. This general notion was fleshed out in more detail in the Dutch study by Eefting (1991) where the well-established thema-rhema structure (or "given" - "new" information) was related to speech tempo. However, the effect of information value was only significant in those cases where additional accentedness was present, which is often, but not always the case in her data. A further link between information struc-

ture and tempo has been mentioned by Klatt (1976) by signalling contrastive information and/or emphasis by slowing down.

Tempo variation on the axis of hyper- and hypospeech

The communication situations listed above made it clear that tempo plays a crucial role in spoken communication which is always in a balance between the economic use of speech production and achieving the communicative goal of being understood by the listener. Using Lindblom's (1990) image that the speech production process changes continually on the hyper- and hypoarticulation axis. Related to synthetic speech production, the task is to find or to model an appropriate balance of the hyper- and hypo-continuum. Consider the two general goals of improving the acceptance of synthetic speech: intelligibility and naturalness. Intelligibility is usually increased by modelling clear speech, i.e. considering hyperarticulation. Improving naturalness is usually achieved by mapping features of conversational speech, i.e. by hypoarticulation. Both ends of this axis have their correlates to speech tempo. Thus, to improve the performance of synthetic speech, both ends of this axis must be considered, and this is dependends on the communicative situation in which a human listener is faced with synthetic speech.

Summary and discussion of chapter 2

This chapter presented a discussion of the most important factors underlying speech-tempo differences that have received attention in the literature (see figure 2.1). The amount of attention devoted to each factor varies considerably, and the factors addressed are presumably not the only ones that operate during speech production. The examples illustrate the great range of situations and conditions in which a change in speech tempo can take place. Since speech unfolds in time there can be no speech without speaking tempo. The tempo of speech is always changing, whether we are aware of it or not. This fact is rarely considered in speech analyses or speech applications (e.g. in speech synthesis where usually only *one* tempo is used, and, presumably, expected to fit all speakers, listeners, situations and text styles).

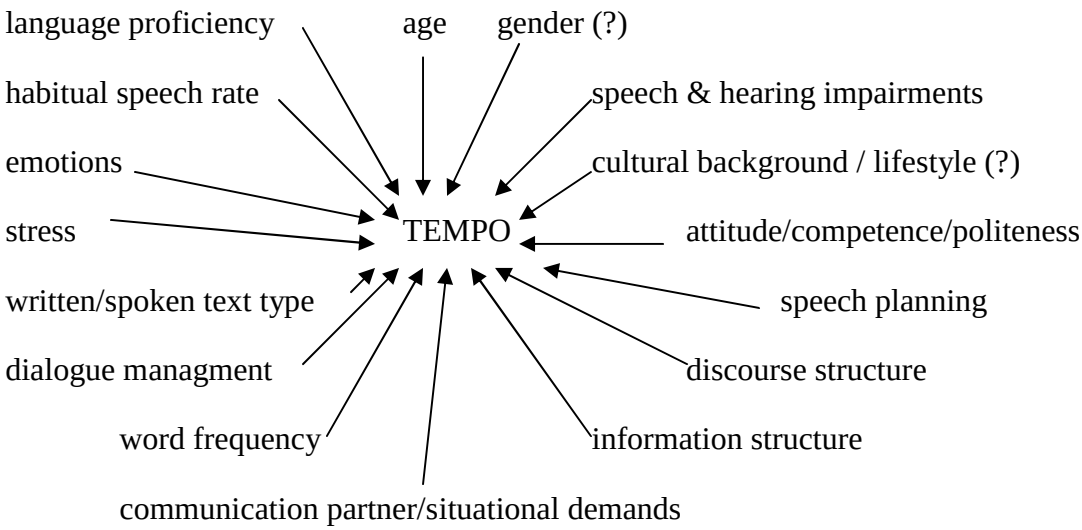


Figure 2.2. Sources of tempo variation.