PHONOLOGICAL ASPECTS OF READING RATE STRATEGIES

Jürgen Trouvain

Abstract

This paper deals with the effect of tempo on phonological structure. In two production experiments, German speakers were asked to read texts at three self-selected rates, "normal", "fast", and "slow". Different speaker strategies were inspected in terms of pausing, phrasing, pitch accent structure and segmental reductions. The first aim of the study is to describe the reorganisation of the phonological structure as a function of the three speech rate categories. The second goal is to discuss the strategies used in speaking faster and slower than normal, considering in particular the homogeneity among speakers and the symmetry within speakers. The differences found between and within speakers provide a basis for modelling individual tempo profiles at the phonological level, which could be exploited e.g. for the synthesis of individual voices and speaking styles.

Dieser Artikel befaßt sich mit der Auswirkung von Sprechtempo auf die phonologische Struktur. In zwei Sprachproduktionsexperimenten wurden Muttersprachler des Deutschen gebeten, Texte in drei verschiedenen Geschwindigkeiten vorzulesen ("normal", "schnell", "langsam"). Die verschiedenen Sprecherstrategien wurden im Hinblick auf Pausierung, Phrasierung, Tonakzentstruktur und segmentale Reduktion untersucht. Das erste Ziel der Studie ist die Beschreibung der Reorganisation der phonologischen Struktur als eine Funktion der drei Tempokategorien. Das zweite Ziel besteht in der Besprechung der zum Schneller- und Langsamersprechen benutzten Strategien unter Berücksichtigung der Homogenität unter den Sprechern und der Symmetrie innerhalb der Sprecher. Die gefundenen Unterschiede zwischen und innerhalb von Sprechern bieten eine Basis zur Modellierung individueller Tempoprofile auf phonologischer Ebene, wie sie z.B. für die Synthese individueller Stimmen und bestimmter Sprechstile genutzt werden können.

Phonus 4, Institute of Phonetics, University of the Saarland, 1999, 15-35.

1. Introduction

1.1. *Reported strategies for speeding up and slowing down*

In the literature a number of different strategies for increasing and decreasing speech rate are discussed. The effect of tempo on phonological structure is manifested at several levels. Allegro and lento rules describe changes in the segmental structure; the tonal structure too has been an object of tempo research; a further strand comes with prosodic phrasing, which is closely linked with the presence or absence of pauses. Strictly speaking, pausing is only one phonetic means of marking a prosodic boundary. However, it has long been accepted as playing a major role in tempo change (Goldman Eisler, 1968). Because of this special status of pauses we choose to treat them here as a separate phonological aspect, apart from phrasing, intonation, and segmental structure.

One aim of this study is to get an idea of the complexity of *phonological* levels and to make statements about re-organised phonetic/phonological structures. It is well known that tempo has strong effects on the *phonetic* realisation in terms of duration, spectral properties, energy, and F_0. So the distinction between phonological structure on the one hand and phonetic implementation on the other is made consciously.

1.1.1. *Pausing*

No matter whether one sees the pause from the production point of view as a silent interval, possibly combined with breathing, or from the perceptual perspective as a perceived pause, two mechanisms for changing the tempo by pausing are possible: a) controlling the number of pauses, and b) controlling the duration of pauses.

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. 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.

Even though Goldman Eisler (1968) claims that pausing is *the* variable that determines speech rate, it is clearly true that pausing, although important, is not the one and only factor.

1.1.2. *Phrasing*

There are parallels and, inevitably, links between phrasing and pausing: on the one hand one can think of a reduction (or increase) in the number of prosodic phrases by deleting (or introducing) prosodic phrase boundaries. On the other hand the shortening (or lengthening) of pauses can be interpreted as the demotion (or promotion) of prosodic phrases, so that major (or minor) phrase breaks become minor (or major) ones.

Indeed, Caspers & Van Heuven (1991) found deletions of phrase boundaries in fast Dutch speech. Additionally to boundary deletions, demotions of 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).

1.1.3. *Intonation*

With respect to the phonology of intonation, researchers have observed changes in the complexity of tonal structures. These changes can be expressed as deletions (or insertions) of accents, and/or in the use of simpler (or more complex) pitch accents. Other intonational properties such as the amplitude of pitch movements, the overall pitch range, or the segmental alignment of pitch targets may be considered part of the phonetic shape rather than the phonological structure. They are therefore not considered further here.

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. In the Dutch production data in Caspers & Van Heuven (1991) boundary marking pitch configurations tended to be simplified in fast speech, though speakers did not reduce

the number of pitch accents. In contrast, Fougeron & Jun (1998) report simplifications in the form of non-realisations of the initial high tone in French.

1.1.4. *Segments*

On the segmental level we are faced with many different kinds of "post-lexical" processes. The so-called "Allegro-rules" comprise all kinds of segmental reductions: deletions e.g. in consonant clusters, degemination, reductions, or assimilations. An overview of the potential processes for German is provided by e.g. Kohler (1990).

It is important to note that the domain in which these connected speech processes apply can vary, but they usually occur within the domain of the (graphical) word. In the case of function words this can lead to the deletion of entire words or even word sequences. Kohler's example "*Hast Du einen Moment Zeit?*" illustrates the word crossing character quite impressively. Careful pronunciation of this sentence can be realised as [ηαστH δυ⌉∪αI≈ν↔ν μο⌉∪μΕντH ∪τσαI≈τH], whereas a reduced form could be [ηασπ μ ‾μο∪μΕν ∪τσαI≈τ]. The only thing that remains from the two words *Du einen* is an assimilated syllabic [μ ‾] that merges into the following initial [μ] of "*Moment*".

Frequently the terms slow and careful, or fast and reduced (or sloppy) respectively, are used synonymously. However, both a fast and careful pronunciation as well as a slow and sloppy style are theoretically possible even though not very probable. Usually, the slow/careful pronunciation is equal to a sequence of words in their canonical forms, and the fast/sloppy end of this scale is used to demonstrate the reduced forms.

1.2. ***Homogeneity and symmetry of strategies***

Considering only the slow rate in the light of the strategies observed for the fast rate, the most simple assumption about what happens would be to assume the reverse of the speeding up strategies. With Goldman Eisler (1968) in mind, one might naively assume that slowing down is always marked with more and longer pauses, and speeding up shows fewer and shorter pauses. This we call the "symmetry assumption".

Furthermore, if one observes that all speakers do more or less the same when speaking faster, one would presume that they will do so also for speaking slowly. This is the "homogeneity assumption".

The examples in Butcher (1981) and Fletcher (1987) provide evidence against the homogeneity and symmetry hypotheses, while Grosjean (1979: 379) found a pattern that supports both ideas: "speakers primarily alter the number of pauses and leave their pause durations relatively constant."

In view of the variety of phenomena that have been observed and the differences between researchers' observations, we shall investigate the hypothesis that not every speaker makes use of all possibilities in the same way and to the same extent.

1.3. *Aims of the Study*

Most speech rate studies work with read material. It is known that non-planned or spontaneous speech is performed differently in many (mainly prosodic) respects (see e.g. Blaauw, 1995). Dysfluencies and ungrammatical constructions which spontaneous speech is notorious for (syllable drawls, filled pauses, false starts, etc.) heavily influence the temporal and also the phonological structure of what is being said. Additionally there are clear hints that global tempo is changed quite dynamically in the course of a conversation (see e.g. Koopmans-Van Beinum & Van Donzel, 1996; or Uhmann, 1989).

In order to make the difference clear between read and spontaneous speech, where we would expect additional and special problems, we use the term *reading rate* as a more specific form of the rather general notion *speech rate.*

We report two production experiments here in which we asked people to read texts of paragraph length at three different rates "medium", "slow", and "fast". In both experiments we ascertained speaking and articulation rate, mean pause duration as well as the number of pauses and the number of prosodic phrases. In experiment 1 we focus on pausing structure, phrase structure and segmental reductions, whereas in the second experiment (which has also been reported in Trouvain & Grice, 1999) the focus lies on phrasing and tonal structure.

Apart from the question whether German speakers make use of the various possibilities for phonological rearrangement (see section 1.1), the analysis and the

interpretation of the results are discussed against the background of the homogeneity and the symmetry assumptions versus individual strategies.

We summarise the questions as follows:

- Are the strategies at the various levels used by all speakers?
- Are the strategies that are used for speeding up also used to a similar extent for slowing down?
- Is it possible to sketch an individual tempo profile at the phonological level?

2. Experiment 1

2.1. *Methods*

In the first experiment three female native speakers of German recorded three readings of a five-sentence newspaper article (see Appendix for text) at three self-selected rates, "normal", "fast", and "slow". This resulted in 27 versions for analysis (3 speakers x 3 rates x 3 readings). Each cycle of readings started with "normal" followed by "slow", and finished with "fast". The speakers (labelled *speaker 1, 2, 3*) come from different dialect regions (Moselfränkisch, Badisch, Schwäbisch). Although no one showed obvious dialectal peculiarities at the segmental level, regional influences cannot be completely excluded.

For each version the total duration (in ms) was measured as well as the durations of pauses longer than 100 ms (cf. Butcher 1981), extended to 150 ms when the pause was followed by a stop consonant. Closure durations in post-pausal positions were counted as part of the total pausing time, however. The speaking rate (*including* pauses) and the articulation rate (*excluding* pauses) are calculated as a function of the number of phonological syllables (which is same for all versions).

For the second of the five sentences in the text (arbitrarily chosen), phrase breaks were transcribed impressionistically by the author, allowing for a three-level distinction (0 = no break, 1 = minor break, 2 = major break).

As an illustration of segmental processes, a transcription of a short excerpt from the first sentence (consisting of 4 words) is given.

2.2. *Results*

2.2.1. *Speaking rate and articulation rate*

In neither representation of the global rate (see table 1) is there any overlap of the three tempo categories between speakers. That means that, across our subjects, the realised rate categories correspond to the intended ones.

Table 1. Mean values (in phonological syllables per second) for speaking rate (SR) and articulation rate (AR) for the three speakers for each of the intended rate.

	SR			**AR**		
Speaker	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Fast	5.54	6.12	6.75	6.06	6.49	7.43
Normal	4.33	4.84	4.98	4.81	5.42	5.68
Slow	3.44	3.80	3.55	3.90	4.49	4.13

The differences between the speakers are seen in the mean values for their normal tempo and in the values expressing their tempo range. Speaker 3 prefers a higher speed than the others, and she also shows a wider range.

2.2.2. *Pausing*

All three subjects have a greater increase in the number of pauses from "normal" to "slow" than from "fast" to "normal". The greatest increase is found for speaker 2, followed by speaker 3, and then by speaker 1 (see table 2).

All speakers also show a homogenous picture in terms of mean pause duration: the slower they speak, the longer the pauses. Nonetheless, the differences for slowing down are smaller than those for speeding up.

Table 2. Distribution of all realised pauses in 100 ms bins for each of the three speakers for each rate (slow, normal, fast). Maxima are in bold (see text). Mean number of pauses and the mean pause durations are given in the last two rows.

Speaker	**S1**			**S2**			**S3**		
Rate	**S**	**N**	**F**	**S**	**N**	**F**	**S**	**N**	**F**
100-200	1	-	-	**8**	1	2	2	-	-
200-300	-	-	-	2	-	1	-	-	1
300-400	2	-	**6**	**4**	1	**5**	3	-	**5**
400-500	1	**4**	2	2	**5**	1	2	1	**4**
500-600	1	2	3	**4**	**6**	-	-	**4**	-
600-700	**6**	**4**	-	**5**	2	-	2	**5**	-
700-800	**5**	1	-	2	-	-	2	1	-
800-900	-	1	-	2	-	-	**5**	1	-
> 900	1	-	-	-	-	-	3	-	-
mean no.	5.3	4.0	3.7	9.7	4.7	4.0	6.3	4.0	3.7
mean dur.	666	594	428	438	424	258	664	635	379

Pauses show a great diversity in temporal extension. The distributions of pauses in bins differing in steps of 100 ms duration can illustrate some of the regularities in pause timing. In the "fast" condition all speakers produce most of their pauses in the duration group between 300 and 400 ms, with a slight tendency to shorter durations for speaker 2, and a tendency to longer durations for the other two speakers.

This tendency is continued in the "normal" pauses where speaker 3 has her maxima between 500 and 700 ms, speaker 2 between 400 and 600 ms, and there is a "bimodal" distribution for speaker 1 (400-500 ms, and 600-700 ms). Such a division is also present in the "slow" data for speaker 2: although she also uses the whole range of durations, she structures her pause durations in long (500-700 ms), medium (300-400 ms) and short (100-200 ms) pauses. Regarding the "short" pause, it was striking that sometimes silent intervals below the 100 ms threshold could be observed for this speaker, though they were not taken into account.

The other two speakers also make use of the entire durational spectrum with higher value maxima, speaker 1 between 600 and 800 ms, and speaker 3 between 800 and 900 ms.

2.2.3. *Phrasing*

Changes in phrasing are illustrated with one example sentence. Table 3 shows the mean break strength for each potential prosodic boundary of the second sentence: *Nach Auskunft* (A) *der Polizei* (B) *war der Junge* (C) *bei einer Klettertour* (D) *an einem Steilhang* (E) *ausgerutscht.* (F) *Im Fall ...*

Table 3. Means of transcribed break strength (0 = no break, 1 = minor break, 2 = major break) for each potential break location. Breakdown per rate (slow, normal, fast) and for each speaker (S1, S2, S3). Numbers after the slash give the mean pause duration (in ms). If no mean pause duration is indicated, no pause was performed.

	S1			**S2**			**S3**		
	slow	**normal**	**fast**	**slow**	**normal**	**fast**	**slow**	**normal**	**fast**
A	0.7	0.3	-	0.7	0.7	-	0.7	-	-
B	1.7/143	2.0	0.7	2.0/176	1.7	0.3	1.7/124	1.3	0.3
C	1.0	0.7	-	1.0	1.3	-	1.0	-	-
D	1.0	1.0	-	1.0/171	0.3	-	1.7/114	0.3	0.3
E	1.0/038	0.3	-	0.3/071	0.3	-	1.3	-	-
F	2.0/740	2.0/609	2.0/448	2.0/687	2.0/531	1.7/140	2.0/910	2.0/665	2.0/346

It can easily be seen that the principle generally holds: the faster the rate, the lower the break level. Exceptions are two cases where "normal" has a slightly higher mean level than "slow" (S1, B, and S2, C), and the end-of-sentence break (F) which remains constant (except for S2 "fast").

As expected, the pause duration for breaks of the same level decreases with increasing rate (break F). However, we can see different pause durations for comparable boundaries not only across rates, but also within a rate category. The pause durations associated with the two transcribed major breaks of the "slow" versions of speaker 2 differ considerably (176 ms vs. 687 ms). The same is true for speaker 1. Her major breaks for "normal" speaking rate are realised at location F with a rather long pause, and at location B with no pause at all. Further examples can also be found for minor breaks.

2.2.4. *Segmental reductions*

Segmental reduction processes are shown in the following example. The word sequence *hat am Morgen einen* from the first sentence (word-by-word translation: a 16 year old *has in the morning an* 80 meter fall ... survived) has as its phonemic form:

/ η α τ α μ μ ρ √ γ ↔ ν α I √ ν ↔ ν /

After the application of several realisation rules such as aspiration of fortis stops, glottal stop insertion before vowels, degemination, and r-vocalisation, one could predict the following phonetic form for clear and slow speech:

[η α τH ? α μ ≈ γ ↔ ν ? αI≈ ν ↔ ν]

One location within this word sequence was selected to describe various phonological processes of connected speech that can apply as reductions from the predicted form. In this example we consider the phoneme sequence at the boundary between the words *ha**t a**m*. Three different processes can occur:

- Omission (or non-realisation) of aspiration [η α τH] -> [η α τ]
- Omission (or non-insertion) of glottal stop [? α μ] -> [α μ]
- Lenition of fortis plosive (with omission/non-insertion of glottal stop) [η α τH ? α μ] -> [η α δ α μ]

Four different versions of the realisation of this bi-phonemic combination are possible. The level of reduction can be expressed by the number of missing or changed phonetic segments compared to the predicted slow/clear form: level 0: [τ H ? α]; level 1: [τ ? α] or [τ H α]; level 2: [τ α], level 3: [δ α].

In table 4 one can see that all possibilities are indeed used and that the speakers reduce more at higher rates. An exception in this respect is speaker 1 who always uses the same forms for "normal" and "fast". Apart from this exception, there is no case where a speaker always uses one form for one rate category. This is particularly true for the slow versions, where we would expect a careful and precise articulation. But only three out of nine realisations correspond to the predicted slow/clear form.

Several reduced forms are chosen for more than one rate, e.g. [τ α] for all speakers, and in case of speaker 3 for all rates. So, a specific reduced form seems not to be bound to a specific rate, and conversely a specific rate seems not to be bound to a specific reduced form and the processes which generates it. This is shown in the fact that speaker 1 is the only one not to use different forms for any given rate.

Table 4. Frequencies of realisations of "ha**t** **a**m Morgen". The level of reduction (from 0 to 3) depends on the number of the phonological processes applied to a quasi-canonical form. Each speaker (S1, S2, S3) produced 3 versions at each rate (slow, normal, fast).

		S1			**S2**			**S3**		
reduction		**S**	**N**	**F**	**S**	**N**	**F**	**S**	**N**	**F**
0	τ H ? α	1	-	-	2	-	-	-	-	-
1	τ ? α	2	-	-	-	2	-	2	1	-
1	τ H α	-	-	-	1	-	1	-	-	-
2	τ α	-	3	3	-	1	2	1	1	1
3	δ α	-	-	-	-	-	-	-	1	2

From the individuality point of view it can be said that speaker 3 has the strongest propensity for reduction in this example, followed by speaker 1, and speaker 2 with the least reduced productions.

A similar pattern can be reported from another example in *Mor**gen ei**nen*. Here, four processes can be employed: omission (or non-realisation) of glottal stop, schwa elision, place assimilation of nasal, and deletion of homorganic plosive. These processes result in six possible forms ranging from [γ↔ν?αI≈] > [γ↔ναI≈] / [γN?αI≈] > [γNαI≈] / [N?αI≈] > to [NαI≈]. Each speaker used the most careful pronunciation only once and the degree of reduction increases gradually across the rates, but the type of reduction is not confined to one tempo category.

2.3. *Summary Experiment 1*

In general, the assumptions about the phonological mechanisms of speeding up and slowing down were confirmed in the experiment. After making sure that the speakers were able to match the intended rate categories "fast", "normal" and "slow" temporally, it has been shown that pause timing works as expected: the slower the speed, the longer and the more frequent the pauses, and vice versa. Both pausing features become evident in the mean values as well as in the overview with the temporal distribution of pauses (table 2).

Similarly in the case of phrasing, which is illustrated on the basis of one example sentence: the slower (or faster) the speaking rate, the more (or less) prosodic breaks and the higher (or lower) the break level.

For the segmental reduction phenomena, too, the expectations were fulfilled on a general level: the faster the speech, the more reduced forms are selected.

But these observations can neither express the degree of generalisation nor individual tendencies. Even if we can say something general about rate and reduced forms, that does not mean that at a slow rate *in general*, i.e. in the majority of the cases, the most precise form is produced, and for fast rate the most reduced one. There seems to be a lot of optionality, which sometimes forms individual patterns such as the three-fold distinction of the pauses for slow speech of speaker 2.

Another example of the restricted value of a general statement is the assumption that higher level boundaries are associated with longer pauses, and lower level boundaries with shorter or even absence of pauses. This general hypothesis was confirmed here. Nevertheless, pause duration is not tied to a certain break level: speakers select long pause durations as well as short pause durations for the same break level within a certain rate, and they make differences in pause duration for the same break level across rates. A more differentiated analysis of the break levels, as in De Pijper & Sanderman (1994), might lead to a better correlation between break level and pause duration.

3. Experiment 2

3.1. *Methods*

For the second experiment, three female native speakers of German recorded three readings of the German version of "The North Wind and the Sun" (see Appendix for text) at three self-selected rates, "normal", "fast", and "slow". This resulted in 27 versions for analysis (3 speakers x 3 rates x 3 readings). The experiment is described in detail in Trouvain & Grice (1999).

The procedure is the same as for experiment 1 with the following exceptions: the phrase breaks are transcribed by two labellers (one was the author). To illustrate the change of boundary strength an index with three levels was defined for each reading: a shift from major to minor boundary would involve a -1 step, a shift from no to a major boundary is +2, and so on. All steps are summed to give a cumulative shift

value. Furthermore, transcribed pitch accents are divided into two groups, bitonal and monotonal ones.

All speakers (labelled *speaker 1, 2, 3*) stem from southwest German dialect regions (Saarbrücken (=Rheinfränkisch) & Badisch). Again, no one showed obvious dialectal peculiarities at the segmental level. Regional influences, especially in intonation, cannot be excluded. Speaker 2 also participated in the previous experiment.

3.2. *Results*

3.2.1. *Speaking and articulation rate*

Table 5 shows the results for the rate characteristics which are similar for both measurements, speaking rate and articulation rate: speakers 1 and 2 make clear differences between the three rates whereas for speaker 3 the difference between "fast" and "normal" is only small.

Unlike experiment 1, the speakers in this experiment do not form a homogenous group with regard to their speech rate categories. Speaker 3 is generally slower for all three categories and also has a smaller range between "slow" and "fast". Her "fast" category is almost as slow as the "slow" versions of the other two speakers.

Table 5. Mean values (in phonological syllables per second) for speaking rate (SR) and articulation rate (AR) for the three speakers for each of the intended rate.

	SR			**AR**		
Speaker	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Fast	5.51	6.27	4.39	6.33	6.93	5.19
Normal	4.84	4.91	4.26	5.60	5.85	5.11
Slow	4.30	4.14	3.48	5.14	5.08	4.58

3.2.2. *Pausing*

No difference is observed in the number of pauses for speaker 3 between "fast" and "normal" (see table 6).

Table 6. Mean number of pauses and mean pause duration in ms.

	number			**duration**		
Speaker	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Fast	6.7	6.0	13.3	646	465	475
Normal	8.7	11.3	13.0	592	533	548
Slow	11.7	13.3	17.0	583	608	772

However, with pause duration a different picture emerges (also in table 6). All three speakers distinguish the three rates in terms of pause duration, though not in the same way. Speakers 3 and 2 both increase pause duration as rate decreases while speaker 1 does the opposite, lengthening the average pause duration as she increases the rate, though she reduces the number.

3.2.3. *Phrasing*

Regarding the number of transcribed breaks (see table 7) speaker 1 & 2 make distinctions between the three rates, although they do this to a different extent. Speaker 3 again makes no distinction between "fast" and "normal", but we can see a clear difference between "slow" and "normal".

Table 7. Mean number of transcribed prosodic breaks.

	breaks		
Speaker	**S1**	**S2**	**S3**
Fast	18.0	15.3	19.7
Normal	19.7	18.7	20.3
Slow	20.7	21.3	26.4

The summing of all boundary strength steps shows that speaker 2 demotes phrases for speeding up and promotes phrases for slowing down (figure 1). Speaker 1 only applies demotion for speeding up, and speaker 3 only uses promotion for slowing down (the sum of the break indices are equal for "fast" and "normal").

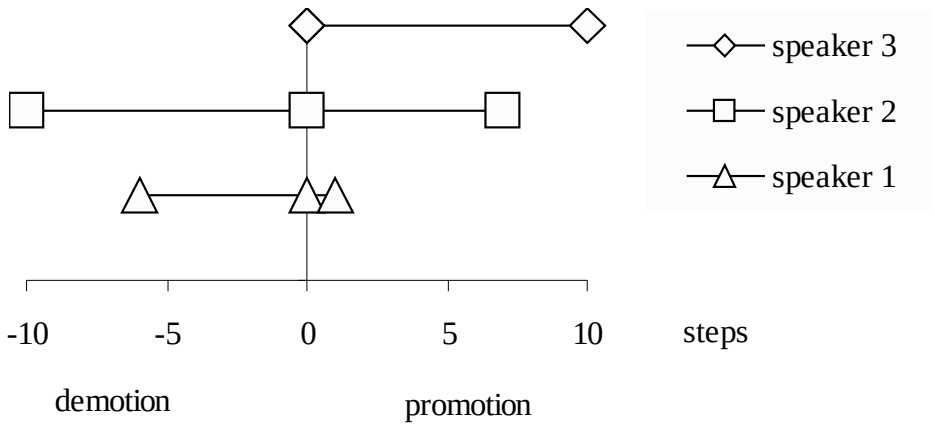

Figure 1. "Promotion" and "demotion" of prosodic boundaries taken for each speaker separately, comparing normal rate to fast and normal to slow. Break index score changes are calculated in steps.

3.2.4. *Pitch accents*

As stated in section 1.1., pitch accents can be expected to be reduced in number and complexity. The results of the pitch accent analysis show considerable variation (see figure 2). Speaker 1 has almost no change from "normal" to "slow" and no change in the total number of accents from "normal" to "fast", but a considerable reduction in the number of bitonal accents. Speaker 2 increases and decreases, respectively, both the total number of accents and the number and proportions of bitonal accents from "normal" to "slow" and from "normal" to "fast". Speaker 3 shows the same pattern of increase in total accent number, but there is no change in the number of bitonal accents with rate, resulting in a reverse pattern in the proportion of bitonal accents fast > normal > slow.

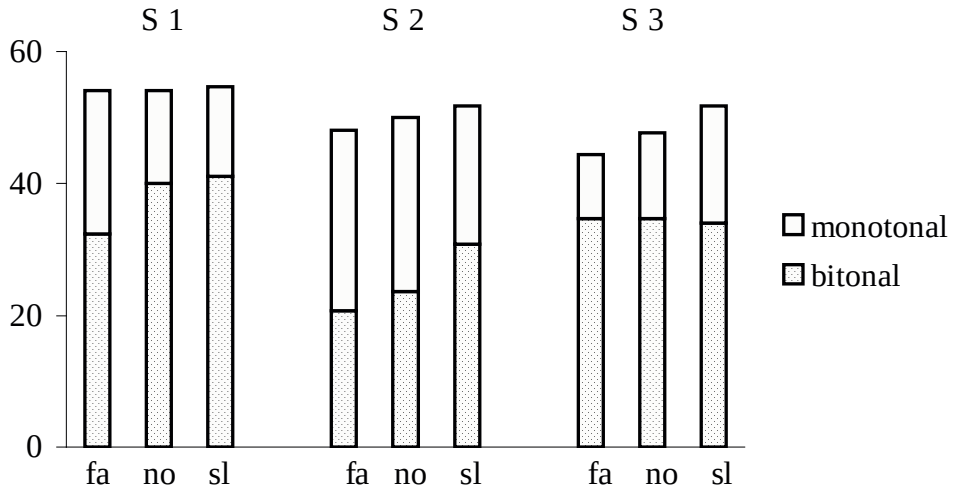

Figure 2. Distribution of pitch accents for all realisations. Pitch accents are classified either as monotonal or bitonal.

3.3. *Summary Experiment 2*

In the second experiment we checked the consistency of the results from the first experiment with respect to articulation and speaking rate as well as mean number and mean duration of pauses. Although the same general patterns were found in experiment 2, there are some interesting differences. Speakers differ more strongly in their choice of strategies, and also in the implementation of these strategies. Even within one speaker (speaker 2 participated in both experiments) do we find different magnitudes for the examined parameters.

In experiment 2 phrasing was investigated in terms of number of breaks and the change of break level, whereas in experiment 1 the break locations and their reflection in pause duration was explored. Additionally, the tonal pattern was analysed in the second experiment.

The second experiment reveals individual patterns of change rather than general tendencies such as those in the first experiment. Speaker 2 conforms to the patterns of change regarding the number of pauses, the number of phrases, the number of pitch accents (especially the number of bitonal ones), the pause duration, and the promotion and demotion of phrase boundaries, that are expected on theoretical grounds. The other two are inconsistent in several of the analysed properties: Speaker 1 with respect to a) mean pause duration, b) promotion of phrases, and c) number of pitch accents.

Speaker 3 concerning a) articulation and speaking rate, b) mean pause duration, and c) demotion of phrases.

4. Discussion

4.1. *Phonological reorganisation*

In a recent article Ladd et al. (1999) criticise the rather general observations in investigations of the effects of speech rate on intonation:

> "Relative number of prosodic boundaries and relative F0 level are global properties of contours, and it is therefore difficult to incorporate the findings (...) into a quantitative model. In particular, knowing about such global effects is of little or no use for predicting or modeling the effect of changes in speech rate on the detailed course of F0 in individual pitch accents."

The same criticism can be applied to the more global analysis of this study. But it is necessary to know the overall patterns of change before details make sense. Even though global statements about changing phonological properties cannot predict the final shape of phonetic parameters, they can help to model the changed phonological frame which forms the basis for predicting phonetic parameter values.

It is evident that speech rate affects the phonological structure in such a way that on this basis alone segment and syllable durations are changed: de-accentuation results in a lack of accentual lengthening; the promotion (or demotion) of a phrase boundary results in increased (or reduced) phrase-final lengthening; vowel reduction can lead to a vowel type which is reduced in its inherent duration; a degeminated consonant is shorter than two consonants, and a deleted consonant means zero duration. So, the starting-point for predicting or explaining segment durations is highly dependent on the reorganised phonological facts.

In the case of spectral reduction we can see that a (quite complex) re-structuring of the phonological frame ultimately linked with tempo change needs to be considered when purely phonetic properties such as spectral quality are investigated. Fast speech alone need not trigger spectral reduction, but it usually occurs together with other

prosodic conditions like accentedness, or position in a prosodic phrase. These conditions are affected by tempo, however: the degree of accent can be reduced, or the length (as well as the duration) of a prosodic phrase can change.

Knowing about re-structuring of pitch accentuation is important, whether or not we assume that different underlying tones were realised, or think that monotonal pitch accents are reduced bitonal ones.

The problem with global statements of the kind cited above is that a) they only show general tendencies, and do not capture the strategies applied by a real speaker, and b) they say something of *which* variables change, sometimes a little bit of *how* variables change, but they usually say nothing about *when and where* variables change.

This study attempts to generalise in various respects. It looks at slow *and* normal *and* fast speech, not only the fast-normal distinction, though the author is aware of the fact that these tempo categories are artificial. It looks at several segmental and suprasegmental phenomena, not only one aspect of phonology. But it also looks behind the scene of the general tendencies and tries to shed some light onto the mechanisms for achieving tempo change.

What the study does not do is to make clear the *when and where*, i.e. under which conditions exactly a modification rule is applied and to what degree. It is good to know, that, let us say, 20% of minor prosodic boundaries should be "demoted" in fast speech, but this says nothing about the exact conditions nor about which breaks are concerned. It is of course necessary to have more insight into the location of boundaries. It is a truism that "more important" boundaries are realised more elaborated (longer pauses, more final lengthening, boundary tones, creaky voice, ...).

4.2. *Individual strategy profiles*

What is clear from this study is that speakers differ in their strategies for achieving another tempo, and that these differences can be quite considerable. It also becomes evident that strategies for slowing down are not reversed speeding up strategies. This lack of homogeneity among speakers and the lack of symmetry within speakers are important features for modelling speech rate, both for a general tempo model, and for an individual model.

Of course, individual strategies have frequently been observed on various phonetic levels. E.g., in the study by Ladd et al. (1999) only certain speakers enlarge

the pitch excursion size as rate slows down, and Kuehn & Moll (1976) report different preferences in terms of velocity and displacement of articulators.

Personality markers are also apparent in spontaneous speech. The type of pause fillings, the way syllables are drawled, the locations of interruptions in the speech flow, and the frequency of all kinds of dysfluencies are substantial features of an idiolect.

These are only three aspects which contribute to an individual tempo profile. Idiosyncrasies need to be considered on many levels. It is likely that strategies of phonological restructuring are important ones. If we think of speech synthesis applications, it is a necessary start. No matter whether one wishes to develop an individual synthetic voice, e.g. for an animated character, or to enhance the variety of speaking styles for different situations or text styles, symbolic (=phonological) input is always required.

Acknowledgements

The author would like to thank Marc Schröder, Bill Barry, and Jacques Koreman for valuable discussions.

References

Blaauw, E. (1995). *On the Perceptual Classification of Spontaneous and Read Speech* (Ph.D. diss., Utrecht University).

Butcher, A. (1981). *Aspects of the speech pause: phonetic correlates and communicative functions* (Arbeitsberichte Phonetik Kiel **15)**.

Fletcher, J. (1987). Some micro and macro effects of tempo change on timing in French. *Linguistics* **25**, 951-967.

Fougeron, C. & Jun, S.-A. (1998). Rate effects on French intonation: prosodic organization and phonetic realization. *J. of Phonetics* **26**, 45-69.

Goldman Eisler, F. (1968). *Psycholinguistics*. London & New York: Academic Press.

Grosjean, F. (1979). A study of timing in a manual and a spoken language: American sign language and English. *J. of Psycholinguistic Research* **8**, 379-405.

Kohler, K.J., (1990). Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In: Hardcastle & Marchal (eds), *Speech Production and Speech Modelling*, 69-92.

Koopmans-Van Beinum, F.J. & Van Donzel, M.E. (1996). Relationship between discourse structure and dynamic speech rate. *Proc. IVth. Int. Conf. On Spoken Lang. Proc., Philadelphia*, 1724-1727.

Kuehn, D.P. & Moll, K.L. (1976). A cineradiographic study of VC and CV articulatory velocities. *J. of Phonetics* **3**, 303-320.

Ladd, D.R., Faulkner, D., Faulkner, H. & Schepman, A. (1999): Constant "segmental anchoring" of F_0 movements under changes in speech rate. *J. Acoust. Soc. Am.* **106**, 1543-1555.

De Pijper, J.R. & Sanderman, A.A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *J. Acoust. Soc. Am.* **96**, 2037-2047.

Rietveld, A.C.M. & Gussenhoven, C. (1987). Perceived speech rate and intonation. *J. of Phonetics* **15**, 273-285.

Trouvain, J. & Grice, M. (1999). The effect of tempo on prosodic structure. *Proc. XIVth Int. Conf. On Spoken Lang. Proc., San Franscisco*, 1067-1070.

Uhmann, S. (1989). On some forms and functions of speech rate changes in everyday conversation. *Working Paper "Kontextualisierung durch Rhythmus und Intonation"* No. **7**, Konstanz.

Appendix

Text Experiment 1

Ein Sechzehnjähriger hat am Morgen einen 80-Meter-Sturz in den Silbersee bei Neunkirchen überlebt. Nach Auskunft der Polizei war der Junge bei einer Klettertour an einem Steilhang ausgerutscht. Im Fall schlug er mit dem Kopf auf und fiel bewußtlos in den See. Ein Camper holte den Jungen aus dem Wasser und alarmierte den Rettungsdienst. Der Sechzehnjährige ist außer Lebensgefahr.

Text Experiment 2

Einst stritten sich Nordwind und Sonne, wer von ihnen beiden wohl der Stärkere wäre, als ein Wanderer, der in einen warmen Mantel gehüllt war, des Weges daherkam. Sie wurden einig, daß derjenige für den Stärkeren gelten sollte, der den Wanderer zwingen würde, seinen Mantel abzunehmen.

Der Nordwind blies mit aller Macht, aber je mehr er blies, desto fester hüllte sich der Wanderer in seinen Mantel ein. Endlich gab der Nordwind seinen Kampf auf. Nun erwärmte die Sonne die Luft mit ihren freundlichen Strahlen, und schon nach wenigen Augenblicken zog der Wanderer seinen Mantel aus. Da mußte der Nordwind zugeben, daß die Sonne von ihnen beiden der Stärkere war.