Chapter 6

Production Experiments: Effects of Tempo on Phonological Structure

Introduction

Usually, natural speech occurs as spontaneous face-to-face dialogue. Reading texts aloud is normally tied to specific sorts of speech, e.g. reading newpapers articles or letters to somebody in the room, reading books aloud to children, many types of broadcasted speech (radio and television news, teleprompted speech, weather forecast), many forms of aesthetic communication (radio play, read novels, recitation of poems), or other forms of public-oriented speeches (press conferences, official speeches, judges' verdicts, prayers, sermons, university lectures). For experimental studies performed under laboratory conditions, read speech has the advantage that different versions of one text are more comparable in contrast to different examples of spontaneous speech. Since one of the aims of the study is to explore implications for *text*-to-speech synthesis systems it seems legitimate to examine the production of read rather than spontaneous speech.

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 speakers make use of the various possibilities for phonological rearrangement, the analysis and the interpretation of the results are discussed against the background of homogeneity and symmetry assumptions versus individual strategies.

We summarise the questions as follows:

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

6.1. Production Experiment 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), a threshold that was 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 the same for all versions). For several reasons it was decided to measure the tempo in phonological syllables rather than in sounds although the discussion in chapter 4 has shown that the sound segment, either as realised phone or as phoneme, seems to be the most favourable unit in tempo measurement. First, there are the advantages of the syllable compared to the sound segment: easier counting, easier definition and higher degree of popularity and therefore a higher degree of comparability across studies. Second, there are the advantages of counting intended forms rather than realised ones in terms of reliable identification and, for the sound segment, acceptable definition.

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.

Results

Speaking rate and articulation rate

In neither representation of the global rate (see table 6.1) 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 6.1. Mean values (in phonological syllables per second) for speaking and articulation rate for the three speakers for each of the intended rate.

	speaking rate			**articulation rate**		
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.

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 6.2).

All speakers also show a homogeneous 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 6.2. Distribution of all realised pauses in 100 ms bins for each of the three speakers for each rate. Maxima are in bold (see text). Mean number of pauses per version (per rate and speaker) and the mean pause durations per version are given in the last two rows.

speaker	**S1**			**S2**			**S3**		
pause dur.	**slow**	**normal**	**fast**	**slow**	**normal**	**fast**	**slow**	**normal**	**fast**
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 like speakers 1 and 3 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 some silent intervals below the 100 ms threshold were 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.

Phrasing

Changes in phrasing are illustrated with one example sentence. Table 6.3 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 ...*

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.

Table 6.3. Means within versions 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 had been produced.

	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

If the values for mean break strength and mean pause durations (table 6.3) are averaged across the three speakers (figure 6.1), the following features of the non-linear nature of changes of speech rate in terms of phrasing and pausing can be noted:

- pause duration of breaks of the same strength and of the same location decreases non-linearly from slow to fast (cf. break F)
- some breaks reduce in strength whereas others keep the same strength from slow to fast (cf. breaks A-E vs. F)
- higher level breaks are not necessarily marked by a pause, whereas lower level breaks can be marked by a pause at the same rate (cf. break B vs. E at slow)
- mean break strength correlates non-lineraly with pause duration (cf. break F vs. B vs. D)

The non-uniform changes of the break strength at the same break location across the rates as well as the considerably different pause duration for the breaks of the same level across the rates demonstrate the non-linearity of how tempo change is achieved with respect to phrasing.

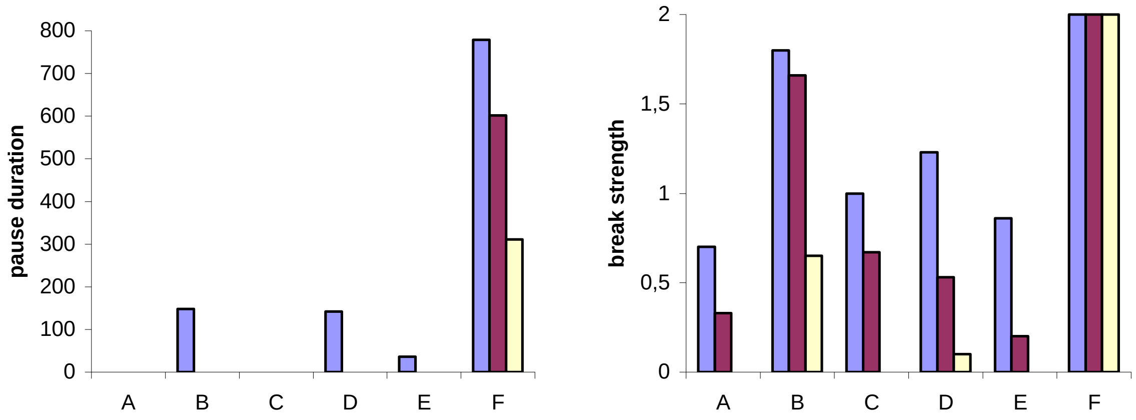

Figure 6.1. Mean pause duration and mean break strength for each break location averaged across the three speakers for each rate category (slow=bright, left; normal=dark, central; fast=white, right).

Segmental reductions

Many reduction processes can be observed in connected speech and they are well described for German (e.g. Kohler, 1990) as well as for other languages as presented in chapter 3. The aim of the analysis in this experiment is to check whether processes apply in a consistent way along speech rates and across speakers. Thus, 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:

/ h a t a m m ɔ r - g ə n aɪ - n ə n /

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:

[h a th ʔ a m ɔɐ g ə n ʔ aɪ n ə n]

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 *hat* ***a****m*. Three different processes can occur:

- Omission (or non-realisation) of aspiration [h a th] -> [h a t]
- Omission (or non-insertion) of glottal stop [ʔ a m] -> [a m]
- Lenition of fortis plosive (with omission/non-insertion of glottal stop) [h a th ʔ a m] -> [h a d a m]

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: [th ʔ a]
- level 1: [t ʔ a] or [th a]
- level 2: [t a]
- level 3: [d a].

In table 6.4 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.

Table 6.4. 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	t h ʔ a	1	-	-	2	-	-	-	-	-
1	t ʔ a	2	-	-	-	2	-	2	1	-
1	t h a	-	-	-	1	-	1	-	-	-
2	t a	-	3	3	-	1	2	1	1	1
3	d a	-	-	-	-	-	-	-	1	2

Several reduced forms are chosen for more than one rate, e.g. [t a] 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 generate it. This is shown in the fact that speaker 1 is the only one not to use different forms for any given rate.

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 the phrase: *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 [gənʔaɪ] > [gənaɪ] / [gŋʔaɪ] > [gŋaɪ] / [ŋʔaɪ] > to [ŋaɪ]. 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.

Summary Experiment 1

In general, the assumptions about the phonological mechanisms of speeding up and slowing down presented in chapter 3 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 6.2).

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 scope for variation, which sometimes results in 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.

6.2. Production Experiment 2

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 boundary 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.

Results

Speaking and articulation rate

Table 6.5 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.

Table 6.5. Mean values (in phonological syllables per second) for speaking and articulation rate for the three speakers for each of the intended rate.

	speaking rate			**articulation rate**		
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

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.

Pausing

No difference is observed in the number of pauses for speaker 3 between "fast" and "normal" (see table 6.6). However, with pause duration a different picture emerges (also in table 6.6). 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.

Table 6.6. Mean number of pauses and mean pause duration in ms.

	number of pauses			**pause 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

Phrasing

Regarding the number of transcribed breaks (see table 6.7) 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 6.7. Mean number of transcribed prosodic breaks.

	number of 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 6.2). 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").

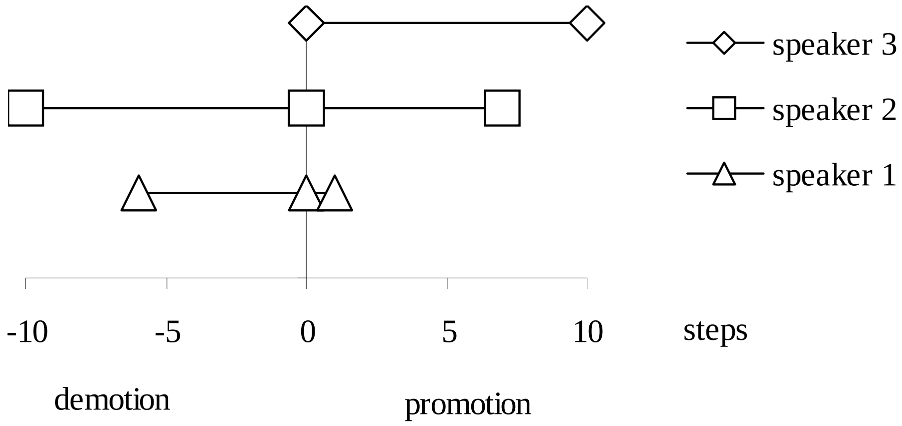

Figure 6.2. "Promotion" and "demotion" of prosodic boundaries taken for each speaker separately, comparing normal rate to fast and normal to slow. Break index score changes are calculated in steps.

Pitch accents

As stated in chapter 3, pitch accents can be expected to be reduced in number and complexity. The results of the pitch accent analysis show considerable variation (see figure 6.3). 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.

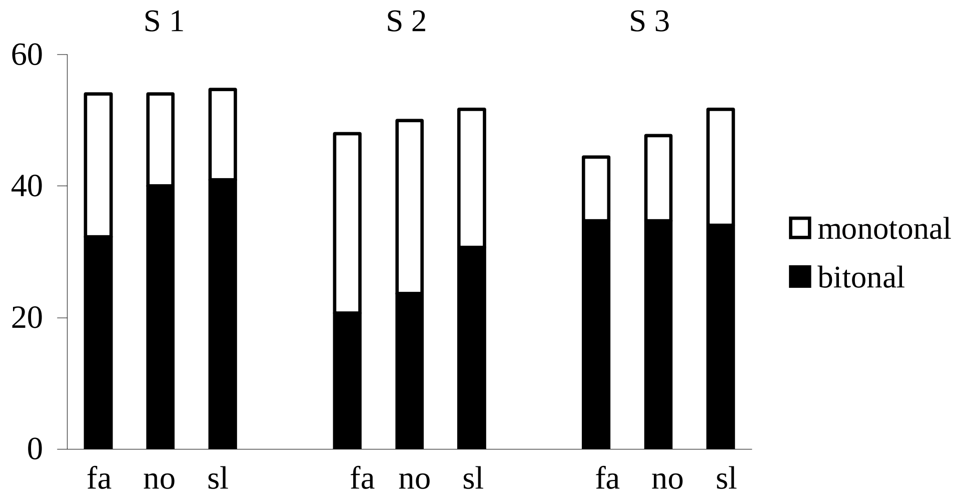

Figure 6.3. Distribution of pitch accents for all realisations. Pitch accents are classified either as monotonal or bitonal.

Summary Production 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 differed 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) we find a change in magnitudes for the examined parameters between experiments.

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.

6.3. Discussion of experiments 1 + 2

Pausing strategies

An expectation from many studies of speech rate and pause behaviour would be that most speakers would make maximal use of pausing mechanisms, namely reducing number and duration of pauses for speeding up, and increasing number and duration of pauses for slowing down. From a brief analysis of pausing strategies across different studies and different languages it becomes clear that this is obviously not the case and that speakers use pausing mechanisms only partially as illustrated in table 3.1 (chapter 3, p. 26). Table 6.8 summarises all the main findings from the above mentioned studies with the result that there is virtually *no* tendency to use *all* possible pausing mechanisms to change tempo. This picture is also mirrored in the strategies of the six speakers described in Trouvain (1999) where only one speaker comes to the maximum exploitation of pausing changes.

Table 6.8: increase (>), decrease (<), or no difference (=) to normal speeded tempo regarding number and duration of pauses.

	slow		fast	
	number	duration	number	duration.
Exp. 1 S1	>	>	=	<
S2	>	=	<	<
S3	>	≥	=	<
Exp. 2 S1	>	=	<	> (sic!)
S2	>	>	<	<
S3	>	>	=	<

Phonological reorganisation

In Ladd et al. (1999) the rather general observations in investigations of the effects of speech rate on intonation were criticised:

> "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 the experiments presented here. But it is necessary to know the overall patterns of change before details can 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 the 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 elaboratedly (longer pauses, more final lengthening, boundary tones, creaky voice, ...)

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 (see figure 6.3). 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.

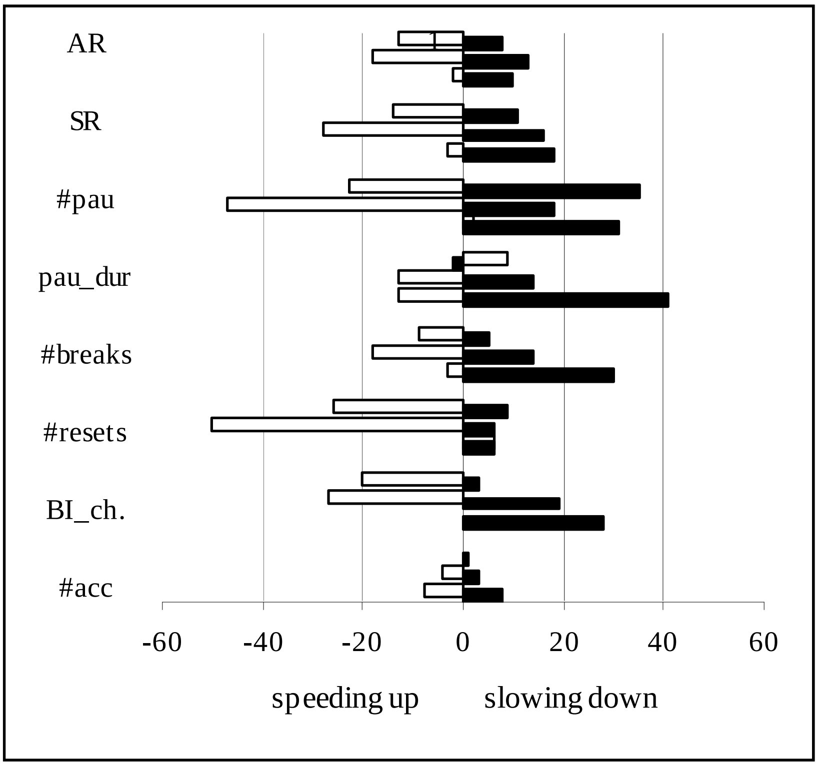

Figure 6.3. Summary of strategies for speeding up and slowing down, expressed as percentages of the normal tempo value for eight parameters. For each parameter, the values for each speaker 1, 2, 3 are given separately: 1. Articulation rate (AR), positive values here indicate fewer syll/s leading to a slower rate, negative values more syll/s; 2. Speech rate (SR), calculated as for AR; 3. Number of pauses (#pau); 4. Average pause duration (pau_dur); 5. Number of transcribed breaks at level 3 or 4 (#breaks); 6. Number of F0 topline resets (#resets); 7. Promotion or demotion of transcribed prosodic breaks (BI_ch.) calculated in steps as for table 6.7; 8. Number of pitch accents (#acc).

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. 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.

Summary and conclusion chapter 6

The production experiments described here confirmed many points which were expected from the knowledge summarised in chapter 3, but they also revealed some details not expected in this form. Some general tendencies for speeding up and slowing down were found to apply, like segmental reduction processes, changing the number of pauses and prosodic breaks, altering the pause durations, and changing number and type of pitch accents. A closer look into individual strategies achieving tempo change shows that this general tendency does not take place for all speakers nor does it occur in the same magnitude nor in each realisation.

With regard to pause durations speakers obviously apply different classes for the three rates. But there is no cue that a general pattern of short, medium and long pauses (cf. Crystal, 1969; Butcher, 1981) holds for one of the speakers in the first experiment. Similarly, the general correlation between break strength and pause duration is questioned. Prosodic breaks of the same strength were marked by considerably higher pause durations when located between sentences compared to breaks located within sentences. Here, syntactic embeddedness plays an important role. Rules that map syntactic breaks to prosodic breaks remain only tendential, while we have to observe that, despite the same syntactic structure, the location of prosodic phrase breaks differ in *all* of the 54 versions in the two experiments. Optionality seems also to play a role in the way segmental structure is re-organised due to tempo change. Rules can describe various processes which are likely to occur but it remains unclear whether, and how systematically these many processes apply. A similar picture emerges with the prosodic processes investigated in the second experiment. Although one of the three speakers could be considered as prototypical, two of them follow their own and sometimes not very consistent ways to vary speech tempo.