The Effect of Tempo on Prosodic Structure

Jürgen Trouvain & Martine Grice
Institute of Phonetics, University of the Saarland, Saarbrücken, Germany

ABSTRACT

This study investigates the effect of tempo on prosodic structuring at both temporal and melodic levels. Readings of a German text are examined to ascertain to what extent changes in pausing, prosodic phrasing, pitch accent structure and F0 realisations contribute towards strategies for speaking at a faster or slower speed than normal. Furthermore, speeding up strategies are compared to those for slowing down, to investigate how far the speakers' behaviour is symmetrical with respect to each of the parameters examined.

1. INTRODUCTION

It has been found that French speakers use a number of strategies for consciously achieving an increase in speech rate [1]. These include a reduction in the number of phrases and the demotion of major to minor phrases, achieved by deleting phrase boundaries or reducing their strength. This prosodic restructuring is reflected in, inter alia, a reduction in the number and mean duration of pauses. Although considerable inter-speaker variability was observed, it was shown that fast speech was largely characterised by a reduction in overall pitch range and in the amplitude of individual rising and falling pitch movements as well as a simplification of the tonal structure, achieved by the non-realisation of underlying tones.

One might expect that consciously slowing down the speech rate would involve the converse of speeding up strategies, viz. an increase in the number of phrases and the promotion of minor to major phrases. One would also expect a concomitant increase in the number and mean duration of pauses. At the melodic level the most simple hypothesis would involve an increase in the global pitch range and in the F0 displacements representing rising and falling pitch events, and a full realisation of the underlying tones.

In this paper we investigate to what extent German speakers make similar adjustments in the temporal structure in order to increase their speech rate, and whether the rate increase is characterised by similar reductions at the melodic level. We then compare strategies used for speaking faster with those employed for speaking at a slower rate than normal, and investigate how far the relation between each of these strategies is symmetrical. Additionally, since German has a richer pitch accent inventory than French, and a higher degree of optionality regarding the presence of pitch accents, we investigate adjustments in type and number of pitch accents when comparing fast and slow rates to a speaker's self-selected normal rate. We do not, however, posit any underlying tones. We take simplification of tonal structure to imply the deletion of one of the tones in a bitonal pitch accent, thus transforming it into a monotonal one, and take the type and number of pitch accents realised to reflect the underlying type and number, since we have no way of determining the presence of an unrealised underlying tone or pitch accent.

2. METHOD

Three female native speakers (CZ, AT and PS) each produced three readings at three different self-selected rates (normal, fast and slow) of the German version of the standard IPA text "The North Wind and the Sun" (Nordwind und Sonne). Pitch accents and boundary tones were transcribed by both authors, using a skeleton GToBI [2], [3] system. Since downstep was not marked in the transcription, the number of pitch accent categories was reduced from 6 to 5 (L+H*, L*+H, H*, H+L*, L*). The GToBI system allows for two levels of prosodic phrasing: the intermediate, or minor, phrase and the intonation, or major, phrase. Break indices of 3 and 4 corresponding to the minor and major phrases were inserted automatically. Breaks smaller than 3 were not transcribed and therefore counted as "null boundaries".

Measurements were made for each reading of total speaking time and total pause (silence and breathing) duration. These were used to determine speaking rate (including pauses) and articulation rate (excluding pauses) in syllables per second, where the number of syllables was taken to be the maximum number of syllables contained in the text, rather than the actual number of syllables produced in each reading. We also counted the number of pauses, taking as a lower threshold 100 ms (cf [4]), 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.

In addition, we recorded F0 values at transcribed high and low tones, taking as one category "pitch accent tones" (H or L in L+H*, L*+H, H*, H+L*, L*) together with the high intermediate phrase boundary tone (H-) and, as another category, "all tones". This latter category includes the intermediate phrase L- tone and intonation boundary tone configurations of the type H-H% and L-L%. The sequences H-H% and L-(L%) generally have more extreme values than the accent tones [2], due to upstep in the former and final lowering in the latter. The reasons for including H- in the same category as accents were twofold. First, it is defined as having roughly the same F0 as the previous H pitch accent tone [2]. Second, in cases of L*+H before a phrase boundary H-, it was not possible to distinguish the high F0 target for the pitch accent from the H- peak unless the nuclear accent was sufficiently distant from the phrase boundary. That is, in cases where the nucleus was late in the phrase, we were unable to find evidence of two target points. In these cases we marked one target (+H-) which functioned as both the trailing (+H) tone and the phrasal (H-) tone.

For "pitch accent tones" on the one hand and for "all tones" on the other, we calculated the F0 mean, standard deviation and range (highest H to lowest L), and the F0 mean and standard deviation for L tones and H tones as separate categories. In addition, we calculated the mean pitch excursion in rising accents, that is the distance from L to H in either L*+H or L+H* accents.

Assuming that a sequence of targets for H tones from the "pitch accent" category corresponds to a top line, we recorded the position and number of upward shifts, or resets, in this topline. Resets were recorded at positions where a given H tone of the category "pitch accent" was at least 30 Hz higher than the previous H tone in that category. This means that an intervening H-

H% would not prevent a reset from being recorded, even if the pitch height at the H% boundary was considerably higher than the following accent H tone. We take the topline reset to at least partly reflect the tonal contribution of phrasing as opposed to the temporal contribution reflected by pausing, or to a combination of the two reflected in the break index scores. It is also an objective measure in contrast to the break index scores which are entirely subjective.

Glottalisation was prevalent across all speakers, both at syllable onsets in place of or in addition to a glottal stop, and in phrase-final position. Where glottalisation prevented reliable F0 measurements, mainly at L-L% boundaries, we sought, where feasible, a reliable value as near to the boundary as possible. There were a number of cases where no value was recorded (36 out of 62 L-L% labels for speaker CZ and 10 out of 64 L-L% labels for speaker AT).

3. RESULTS

3.1. Rate characteristics

The three speakers vary in the extent to which their articulation rates differ across the three subjective speeds (see table 1). Speaker CZ's values were the most extreme: relative to her normal speed, she increased her articulation rate by 18% to achieve a fast tempo and decreased it by 13% to produce a slow tempo. Speaker AT's values involved an increase of 13% and decrease of 8%. In both cases the increase in articulation rate was greater than the corresponding decrease. By contrast, PS hardly increased her articulation rate at all for her fast readings. Despite this, her slowing down strategy was comparable to that of the other two speakers. That is, she appeared to adjust her articulation rate to achieve a difference in subjective speed, but only in one direction: for slowing down. However, we take care not to conclude from this global measure that her normal and fast speeds were identical in their temporal structure (see section 3.2 below). The fact that PS's normal rate is within the range of the other two speakers' slow rates shows that she can be characterised as a relatively slow speaker.

	AT		CZ		PS	
	AR	**SR**	**AR**	**SR**	**AR**	**SR**
F	6.33	5.51	6.93	6.27	5.19	4.39
	+13%	+14%	+18%	+28%	+2%	+3%
N	5.60	4.84	5.85	4.91	5.11	4.26
	-8%	-11%	-13%	-16%	-10%	-18%
S	5.14	4.30	5.08	4.14	4.58	3.48

Table 1: Mean articulation rates (AR) and speaking rates (SR) in syll/s; differences compared to normal speed in percentages.

Table 1 also shows that for all speakers there are greater differences amongst the three speeds for speaking rate than for articulation rate. This can be observed in particular in the values for speaker CZ, where fast speech has a speaking rate increase of 28% as opposed to 18% for the articulation rate. Since the only difference between the two rate measures is the presence of silence in the speaking rate, it can be deduced, as expected, that speakers make use of pausing to achieve a tempo change.

3.2. Pausing

An adjustment in total pause duration can be achieved by a difference either in the number of pauses, implying deletion or insertion, or in the duration of pauses which are already present. These two strategies are recorded as averages for each speaker and tempo in table 2. All three speakers increased the number of pauses to slow down, whereas only two of the three, AT and CZ, reduced their number to speed up. The one speaker (PS) who barely distinguished normal from fast speeds in her articulation rate, also failed to reduce the number of pauses. She did, however, reduce their mean duration by 13%. This is evidence that the speaker did distinguish the two speeds although a first glance at the global measures might cause one to have doubted this. In fact, this speaker manipulates pause duration to a much greater extent than the other two speakers, both for speeding up and for slowing down (the latter by 41%).

	AT		CZ		PS	
	mean #	**mean dur**	**mean #**	**mean dur**	**mean #**	**mean dur**
F	6,7	646	6,0	465	13,3	475
	-23%	+9%	-47%	-13%	+2%	-13%
N	8,7	592	11,3	533	13,0	548
	+35%	-2%	+18%	+14%	+31%	+41%
S	11,7	583	13,3	608	17,0	772

Table 2: Mean number of pauses and mean pause duration in ms.

3.3. Prosodic phrasing

The transcription of prosodic breaks is influenced in part by pausing and other durational measures such as phrase-final lengthening, and in part by other factors, especially the tonal structure and related F0 values. In section 3.3.1 we therefore discuss not only whether a prosodic break (of level 3 or 4) was transcribed, but also whether there were upward shifts, or resets, in the top line (pitch accent H tone values). We also report in section 3.3.2 on the strength of the breaks transcribed and how this was affected by tempo.

3.3.1. Transcribed breaks and topline resets Two speakers (AT and CZ) showed a considerable reduction in the number of topline resets (26% and 50%) and a moderate reduction in the number of breaks (9% and 18%) in their fast speech. For their slow speech, the corresponding increase in the number of breaks (5% and 14%) and resets (9% and 6%) was not as extreme. Speaker PS did not reduce the number of resets at all in fast speech, nor did she considerably reduce the number of transcribed breaks (only 3% reduction). However, where she did make a major adjustment was in her slowing down strategy: here she increased the number of transcribed breaks by 30%, although this was not reflected in the number of topline resets.

	AT		CZ		PS	
	breaks	**resets**	**breaks**	**resets**	**breaks**	**resets**
F	18.0	5.6	15.3	2.3	19.7	7.0
	-9%	-26%	-18%	-50%	-3%	+6%
N	19.7	7.6	18.7	4.7	20.3	6.6
	+5%	+9%	+14%	+6%	+30%	+6%
S	20.7	8.3	21.3	5.0	26.4	7.0

Table 3: Mean number of transcribed prosodic breaks and topline resets.

3.3.2. Change of boundary strength For each location at which a 3 or 4 break index was transcribed in at least one reading for a given speaker, we counted the number of cases where there was a break, and if so, whether 3 or 4, and the number of cases where there was no transcribed break (referred to as a null break, since no breaks below the level of 3 were transcribed). Where at a given rate a speaker's three readings differed, we took the majority value. Table 4 shows for each speaker separately how often each type of break index promotion and demotion took place and classifies the promotions and demotions in terms of steps. A two step promotion involves a change from a null boundary to a 4 boundary, whereas a two step demotion involves the converse. A one step promotion involves a change from break index 3 to 4 or null to 3, a one step demotion correspondingly involves a change from 3 to null or 4 to 3. We obtain a rough indication of each speaker's strategy by calculating the sum of the step values. We can see from the table that speakers AT and CZ make use of demotion when speeding up (reducing level 4 breaks to level 3 or deleting a level 3 boundary) whereas PS does not use this at all consistently (having the same number of promotions as demotions when speeding up). PS does, however, promote boundaries (by introducing a level 3 break or changing a 3 into a 4, and, at one location, changing a null break into a 4) when slowing down. Speaker CZ also promotes boundaries when slowing down, although to a lesser extent, changing either a break of level 3 into a 4 or introducing a break of level 4 where there was no break in the normal speed versions. We can deduce from the total number of steps in table 4 that speaker CZ uses promotion and demotion, when slowing down and speeding up respectively, whereas the other two speakers do not use both strategies consistently. On the whole, AT demotes boundaries when speeding up and PS promotes them when slowing down.

From the table we can also observe that a great many breaks of level 4 remain as such across the different speeds. No breaks of level 4 were deleted and at only three locations were level 4 breaks inserted when slowing down in all three readings for each location.

steps	BI change	AT		CZ		PS	
		N>F	N>S	N>F	N>S	N>F	N>S
-2	4⇒Ø	-	-	-	-	-	-
-1	4⇒3	5	1	7	-	1	-
-1	3⇒Ø	1	1	3	-	1	-
0	4⇒4	8	12	7	14	16	17
0	3⇒3	5	4	1	1	1	-
+1	Ø⇒3	-	2	-	-	1	5
+1	3⇒4	-	1	-	3	1	3
+2	Ø⇒4	-	-	-	2	-	1
Σ steps		-6	+1	-10	+7	0	+10

Table 4: Promotion and demotion of prosodic boundaries taken for each speaker separately and comparing normal speed to fast and normal speed to slow. Break index score changes are calculated in steps. Two steps: null break becomes 4 (+2) or vice versa (-2); one step: 3 becomes 4 (+1) or vice versa (-1), or null boundary becomes 3 (+1) or vice versa (-1).

3.4. Pitch accents

As can be seen from table 5, speaker PS uses more accents in her slow speech (8% more than in normal) and fewer in her fast speech (8% fewer than normal). Although the other two speakers do not contradict this tendency, the differences between the three speeds are only slight. In fact, speaker AT has no difference at all in the number of pitch accents between normal and fast rates. Speaker AT has instead a different way of changing her accentual structure: at her fast rate she uses 55% more monotonal accents than at normal speed. That is, she simplifies bitonal pitch accents, turning them into monotonal ones, effectively deleting one of the tones in the accent. However, this does not mean that she employs the converse strategy to slow down. Her slow speech is barely different from her normal speech in terms of accent type. Speaker CZ by contrast transforms monotonal pitch accents into bitonal ones when slowing down, but rarely applies the converse strategy (i.e. deleting one of the accent tones) when speeding up. Speaker PS keeps the number of bitonal accents constant and changes the number of monotonal accents: more (36%) for slowing down and fewer (25%) for speeding up.

		AT		CZ		PS	
		mean #	diff	mean #	diff	mean #	diff
F	bi	32.3	-19%	20.7	-13%	34.7	0%
	mo	21.7	+55%	27.3	+4%	9.7	-25%
	Σ	54	0%	48	-4%	44.3	-8%
N	bi	40		23.7		34.7	
	mo	14		26.3		13	
	Σ	54		50		47.7	
S	bi	41	+3%	30.7	+30%	34	-2%
	mo	13.7	-2%	21	-20%	17.7	+36%
	Σ	54.7	+1%	51.7	+3%	51.7	+8%

Table 5: Number of bitonal and monotonal pitch accents, and number of pitch accents regardless of type, averaged across three readings for each tempo and speaker. Percentages are differences from normal tempo.

3.5. F0 characteristics

It was hypothesized in the introductory discussion that F0 realisations of the tonal accents might be modified in such a way that with increasing and decreasing tempo, global pitch range, and F0 displacement would decrease and increase, respectively. In fact, changes in F0 were just as idiosyncratic as the structural features discussed in 3.1-3.4.

3.5.1. Mean F0 For speakers AT and CZ the differences in mean F0 for all tones across the three speeds were negligible (maximally 2Hz between fast and slow). A two-way ANOVA (tempo x subject) showed no significant effect of tempo. There was, as might be expected, a significant subject effect, but no interaction of subject with tempo.

3.5.2. F0 range For F0 range (between the highest H and lowest L tone value) the results for the three speakers again revealed a lack of systematic behaviour. Taking the range between tonal values from pitch accents only, the following observations were made: Speaker CZ retained a constant range across all three speeds. AT and PS both narrowed their pitch range, but whereas AT did so in her fast speech, PS narrowed her range in her slow speech. The way in which the narrowing was achieved also differed, in that AT achieved the narrowing by raising her lowest L tone values (by 6%), and speaker PS did so by lowering (also by 6%) her highest H tones.

Taking the range between all tones, i.e. including boundary tones, there were again differences between speakers, but not necessarily with the same pattern as for pitch accent tones alone: Speaker AT kept a constant range across all three speeds, CZ narrowed her range in fast speech by both raising L and lowering H tones (5% and 3% respectively), and kept the range constant in slow speech by raising both H and L tones (4% and 3% respectively). PS increased her range for slow speech, raising only the H tones (by 7%).

3.5.3. L and H tone values Taking the L and H accents separately, and this time looking at mean values for each tone type rather than the most extreme value for each reading (that is, highest H and lowest L), some statistically significant shifts as a function of tempo were found, though these too were at individual rather than group level. For the L tones, however, there were no significant effects of tempo at all. For the H tones, tempo did not feature as a significant main effect, but interacted significantly with subjects ($F(4,1388)=5.072$, $p<0.001$). A post-hoc comparison of tempo for the individual subjects showed that for CZ there was a monotonic increase in H tone F0 from slow over normal to fast, whereas for PS the lowest F0 was for normal, and the highest for the fast condition, slow being intermediate between the two.

With respect to pitch excursion size in rising (L+H) pitch accents, only two speakers, (AT and CZ) reduced their F0 excursion in their fast speech, albeit to a small degree: by an average of 5 and 4 Hz respectively. These reductions were not, however, significant. There were no differences in excursion size between normal and slow.

4. DISCUSSION

First we discuss differences in the effect of speeding up on the parameters investigated, and how systematic this effect is across speakers: As expected [5], we found a considerable effect of tempo in the pausing structure, either in the number of pauses, as was the case for two of the speakers, or in the mean pause duration, as was the case for the remaining speaker and, to a lesser extent, for one of the other speakers. Our results show that articulation and speaking rate cannot be used as sole indicators of an achieved rate change, since one speaker (PS) showed a negligible difference in these two parameters when speeding up although, as can be seen from figure 1, there were differences in some of the other parameters (notably in pause duration). The effect of tempo on phrasing was systematic for two of the three speakers. This can be seen from the number of transcribed prosodic breaks as well as from changes in boundary strength.

The changes employed when speeding up in the above parameters compare favourably with the results in [1] as far as two of our three speakers are concerned. However, we found little evidence of systematic changes at the melodic level. There were no systematic changes in pitch range, in contrast to [1], which dealt with the same text type but a different language, and to [6] which dealt with isolated sentences but the same language, albeit a different regional variety. There were also only slight tempo-related differences in rising displacement. This contrasted with the results in [1], but since our experiment dealt with pitch excursions from low to high in only L+H pitch accents as opposed to rising displacement of pitch accent and boundary tones, the results were not fully comparable. We did, however, find a degree of simplification in the tonal structure in that bitonal accents were changed into monotonal ones (to a considerable extent by speaker AT), although accents were deleted only sporadically, indicating that, as also found for Dutch [7], pitch accent deletion is not a reliable parameter. We also found an effect of speeding up on topline resets, regarded here as purely melodic indicators of phrasing.

Figure 1 illustrates how far our data supports the simple hypothesis that if a given parameter is changed when speeding up, then the reverse change is employed when slowing down, resulting in a symmetrical relationship. This symmetry is achieved in only some parameters by speaker AT and PS, whereas CZ's values are symmetrical for almost all parameters. We might tentatively consider CZ to be a prototypical speaker, because she employs all of the parameters, which are used by at least one of the other speakers.

This study indicates that there are similarities across languages in a number of strategies employed for speaking at a tempo other than normal. However, we have also observed that slowing down strategies are not always the converse of speeding up strategies, and that individual speakers differ considerably in this respect. These factors must be taken into account when modelling speech tempo.

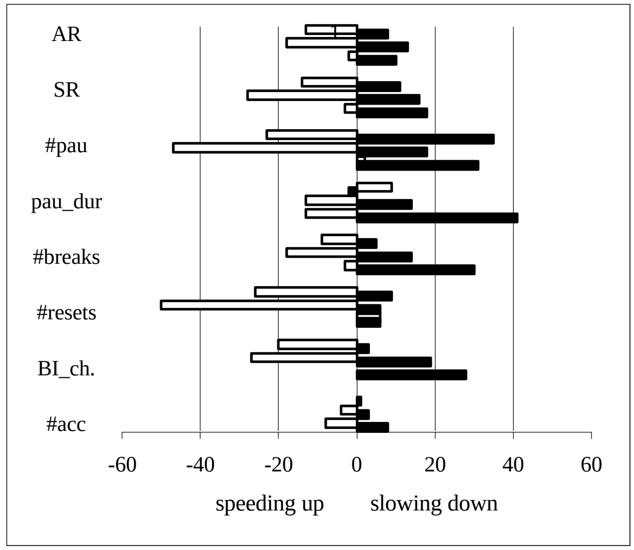

Figure 1: Summary of strategies for speeding up and slowing down, expressed as percentages of the normal tempo value for eight parameters. For each parameter, the values for each speaker are given separately in the following order: AT, CZ, PS. 1. Articulation rate (AR), positive values here indicate 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 4; 8. Number of pitch accents (#acc).

REFERENCES

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

[2] Benzmüller, R. & Grice, M. (1997): Trainingsmaterialien zur Etikettierung deutscher Intonation mit GToBI. Research Report Phonetics, University of the Saarland, Phonus 3, 9-34.

[3] Grice, M., M. Reyelt, R. Benzmüller, J. Mayer & A. Batliner (1996): Consistency in transcription and labelling of German intonation with GToBI. Proc. IV ICSLP, Philadelphia, 1716-1719.

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

[5] Goldman-Eisler, F. (1968): Psycholinguistics. London & New York: Academic Press.

[6] Kohler, K. (1983): F0 in speech timing. Arbeitsberichte Phonetik Kiel 20, 57-97.

[7] Caspers, J. & van Heuven, V. (1991): Phonetic and linguistic aspects of pitch movements in fast speech in Dutch. Proc. XII ICPhS (Vol. 5), Aix-en-Provence, 174-177.