

THE INFLUENCE OF SPEAKERS' OWN SPEECH TEMPO ON THEIR TEMPO PERCEPTION

M. Gósy

Research Institute for Linguistics, Budapest, Hungary.

ABSTRACT

The point of departure for the present paper is the assumption that the speaker's own speech tempo determines his judgements concerning that of other people. Experimental results supported a significant concurrence in tempo perception of 'extreme' speakers as opposed to 'moderate' speakers. A significant correlation was found between the speakers' comprehension and their own speech tempo. It can also be claimed that speakers/listeners judge speech tempo on the basis of the active levels of their speech perception mechanism.

1. INTRODUCTION

Authors of a number of studies agree that tempo perception is basically determined by three factors: articulation rate, the number of pauses, and the duration of pauses [4, 5]. Tempo perception studies are also made difficult by problems concerning the recognition, demonstration, and order of importance of a number of other factors including changes in fundamental frequency (pitch), average intensity, word frequency (of occurrence), syllable structure, rhythmic structure, syntactic properties, etc. [3]. However, very few attempts have been found in the literature to deal with the connection of *production and perception in relation to speech tempo*. It is usually postulated that there should be a very close connection between the speaker's own rate of speaking and his perception and comprehension with respect to speech tempo [1, 2]. Can we then claim that there is a linear connection between tempo production and percep-

tion, namely: the faster the speaker's usual speech tempo the faster his/her comprehension as well? Does this apply to tempo perception, too? What are the criteria of applicability of this rule?

In order to answer these questions, an experiment containing 4 subtests has been performed with Hungarian-speaking native speakers/listeners. The aim of the experiment was to describe the effect of the subjects' own speech tempo on (i) their tempo perception and (ii) their speech comprehension.

2. PROCEDURE

Various methods were used for the subtests. (i) For the first experiment nine speech samples were recorded in random order from Hungarian-speaking native subjects (ages ranged from 25 up to 80). Subjects were selected so that all categories be represented from very slow (articulation rate /AR/: 8.85 sounds/s, overall speech rate /OSR/: 7.25 sounds/s) up to very fast (AR: 18.2 sounds/s, OSR: 14.3 sounds/s). Each speech sample was taken out of a longer monologue and took 1.5 minutes on average. The listeners' task was to judge the speech tempo of each speaker's sample by means of a questionnaire. The categories of the questionnaire were 'very slow', 'slow', 'normal', 'accelerated', 'fast', and 'very fast'.

(ii) The material of the second test consisted of 12 artificial, synthesized sentences (the synthesis was made by a PCF speech synthesizer controlled by an IBM PC). The same sentence had been altered in relation to its overall speech tempo in two ways: by changing

the "articulation rate" of the sentence and by adding one or two pauses at the appropriate grammatical boundary(es) of the sentence. The subjects' task was the same as in the first subtest.

(iii) 8 sentences announced by a trained male speaker were chosen for the third test, and a verification method was used. The sentences were speeded up, and their articulation rate ranged from 20.2 sounds/s up to 24.4 sounds/s. The subjects' task was to decide whether the sentences they heard were true or false. The reaction times (RT) of each subject were measured by means of a fundamental frequency and intensity meter with the accuracy of 10 ms.

(iv) The subjects' spontaneous speech was tape recorded in the final experiment. From their recorded 8-10-minute speech 2-minute samples were picked out for further analysis concerning AR & OSR. The duration and types of pauses were also examined. Counting the speech sounds of the speech sample, the rate was expressed in terms of sound/s.

After finishing the experiments, each subject was asked to judge his/her own average speech tempo according to the formerly used tempo categories. The subjects' sex and age were also recorded on the same answer sheet.

37 subjects were selected from all candidates for further examinations. Three tempo groups were defined: a group of 'slow' speakers, a group of 'moderate' speakers and a group of 'fast' speakers. Examining the data, significant correlation was found between the AR and OSR values of our subjects ($p < 0.05$). 6 subjects were found to be 'fast' speakers in terms of AR and 'moderate' speakers in terms of OSR. So, a fourth tempo category had to be established consisting of subjects having 'fast' AR and 'moderate' OSR and this was labelled the group of 'rapid' speakers.

3. RESULTS

Figure 1 shows the responses of various groups of subjects for all synthesized sentences according to the possible tempo categories. The listeners do

perceive the physical changes of sentences. In the case of sentences containing 1 or 2 pauses, however, the judgements spread along the various tempo categories. The question is whether the distribution of tempo perception is based on the subject's own tempo production. Analyzing the average values for each sentence of each group, it can be stated that there are no important differences among the subjects' judgements. However, the data of the three groups are significantly different at the level of 0.05. This means that there is a slight but definitive difference of tempo perception among subjects with diverse speech tempo production. The mean values of the judgements show very constant changes across the tempo categories. These changes reveal more similarity for the 'slow' and 'fast' speakers than for the 'moderate' and 'rapid' speakers. There is a significant difference in the judgements of the 'slow' speakers concerning the category of 'accelerated' tempo as opposed to the judgements of the other two groups. 'Rapid' speakers' performance shows a relatively different distribution in relation to that of the other two groups. On the basis of these data, a *hypothesis* has been developed on the interrelatedness of the speakers' own tempo production and their tempo perception: 'slow' and 'fast' speakers tend to perceive tempo similarly to one another while 'moderate' speakers do not. 'Rapid' speakers seem to behave perceptually in a way different from the other three groups. We also found that the extreme speakers tend to perceive tempo more on the basis of AR than on the basis of OSR, so the pauses might not influence their tempo perception.

Figure 2 shows the responses of various groups of speakers for the speech samples used across the possible tempo categories. Subjects appear to judge the tempo of the speech samples according to AR rather than on the basis of OSR. The data show: (i) There are larger differences among the tempo categories in each test group than in the case of isolated sentences, and (ii) the distribution of the judgements does not show a constant trend.

The number of responses referring to the 'moderate' tempo category is significantly different in the case of the perception of the synthesized sentences and the speech samples ($p < 0.05$). This means that people's perception mechanism has grown accustomed to the tempo changes of human speech and they are more flexible when judging it than in the case of one sentence where the upper levels of the decoding mechanism should not work, so they can judge the tempo of each sentence more accurate to the actual physical values. The data show again a very similar concurrence of judgements made by the 'slow' and 'fast' speakers. Similar judgements of the 'slow' and 'fast' speakers were found in all tempo categories with the exceptions of the 'fast' and 'very fast' categories. In the case of these two tempo categories the 'slow', the 'moderate' and the 'rapid' speakers judged similarly while the 'fast' speakers differed from all the others. The 'rapid' speakers show a significant difference in their judgements from the other groups of speakers. However, in some cases their judgements fall close to the judgements of one of the groups of speakers. This co-occurrence does not show any systematic character.

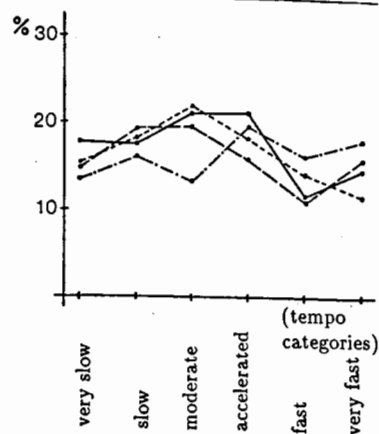


Figure 1.
Tempo perception of sentences by 'slow' (—●—), 'moderate' (---●---), 'fast' (····●····), 'rapid' speakers (—●—●—).

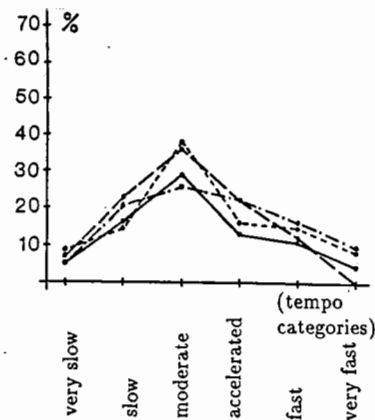


Figure 2.
Tempo perception of texts
see Fig. 1. for the key

The question concerning the analysis of the reaction times was whether the subjects' own speech tempo influences their comprehension rate. The following RT values were obtained: in the case of 'slow' speakers 0.46–1.83 s, in the case of 'moderate' speakers 0.62–1.13 s, in the case of 'fast' speakers 0.3–1.15 s, and in the case of 'rapid' speakers 0.55–0.7 s. A significant difference has been found in reaction times between the various groups of speakers ($p < 0.001$). There is a strong correlation between the subjects' articulation rate and their reaction times which shows that if the tempo of speech production increases the reaction time of the subject decreases. However, there are important differences among the subjects' reaction times within one group. The 'extreme' speakers' reaction times are extreme while the 'moderate' speakers' reaction times are not. There are subjects with fast AR and both with short and long reaction times; and – similarly – there are other subjects with slow AR and with both short and long reaction times. This part of our analysis supports again the similar perceptual behaviour of the 'slow' and 'fast' speakers. The largest RT values were found with the 'slow' speakers and the 'fast' speakers. The 'moderate' and the 'rapid' speakers' RT values were similar

to one another.

There is a significant difference between the RT values of affirmative and negative sentences with true contents; however, there was no significant difference between the same structures with false contents.

Finally, the subjects' age, sex, and their opinion about their own speech tempo were taken into consideration. There was a very strong correlation between the subjects' objectively measured speech tempo and their subjective judgements ($p < 0.001$). We found that most of our extreme speakers were male while the 'moderate' speakers were mainly female subjects ($p < 0.05$). There was no significant correlation between the subjects' age and their speech tempo categories.

4. CONCLUSIONS

– It had been assumed that the faster the speaker's own speech the less fast he perceives that of others. From this hypothesis only the basic point of departure was supported by the results that the speakers' own speech tempo really influences their tempo perception. However, the direction of this influence shows an interesting pattern involving significantly different behaviour for the various groups of speakers. The 'slow' and 'fast' speakers tend to behave perceptually similarly while 'moderate' and 'rapid' speakers tend to differ from the previous two groups. The standard deviation of the reaction time values show the same concurrence for the 'slow' and 'fast' speakers and for the 'moderate' and 'rapid' speakers.

– It has been supported that tempo perception depends primarily on articulation tempo. However, according to our findings, speakers/listeners perceive tempo significantly depending on the activated levels of their whole per-

ception mechanism. If the upper levels of the speech perception mechanism do not play any role in the actual perception process, the tempo judgements (a) are closer to the actual physical parameters of the speech sample and (b) do not show big differences among the speakers having various own speech tempi. If the higher levels also participate in the decisions then other factors (contents of the speech samples, articulation of the speaker, lexicon of the speech sample, timbre, types of hesitations etc.) also play an important role. – On the basis of the significant differences in perception and comprehension of various groups of speakers, we assume that various ways and storage systems should exist for the interactions between the higher and lower levels of the speech perception mechanism determined by the temporal organization of the speakers' speech production.

5. REFERENCES

- [1] BEASLEY, D.S.; MAKI, H.E. (1976), "Time- and frequency-altered speech." In: *Contemporary issues in experimental phonetics*. Ed. LASS, N.J. New York, San Francisco, London, 419-458.
- [2] BOVES, L. (1984), "Perceptual ratings of speech tempo", Dordrecht: Foris
- [3] CUTTING, J.E.; PISONI, D.B. (1978), "An information-processing approach to speech perception", In *Speech and language in the laboratory, school and clinic*. Eds. KAVANAGH, J.F.; STRANGE, W. Cambridge, Massachusetts, London, 38-73.
- [4] FELDSTEIN, S.F.; BOND, R.N. (1981), "Perception of speech rate as a function of vocal intensity and frequency", *Language and Speech*, 24, 387-395.
- [5] den OS, E. (1988), "Rhythm and tempo", Utrecht: Elinkwijk BV.