

ARTICULATORY COMPLEXITY AND THE PERCEPTION OF SPEECH RATE

BERND POMPINO-MARSCHALL HANS G. TILLMANN WOLFGANG GROSSER KARL HUBMAYER

Institut für Phonetik und
Sprachliche Kommunikation
Universität München, FRG

Institut für Anglistik und
Amerikanistik
Universität Salzburg, Austria

ABSTRACT

It is shown that syllable sequences containing complex consonant clusters are perceived as faster than articulatorily less complex ones of the same duration; furthermore, that in AX-discrimination the second test item is perceived as faster.

INTRODUCTION

Although German is supposed to be stress timed, compression of complex stress feet to the duration of simple ones is known not to be complete (2), and thus complex syllables should be perceived as faster in contrast to simpler ones of the same duration (1). With the following experiments we wanted to study this effect in more detail.

METHOD

Three five feet sequences identical with respect to number of syllables, trochaic foot structure, and vowels, but differing in foot initial consonance complexity were uttered by a native speaker of German with the stressed syllables in beat with a computer-generated metronome signal of variable frequency. These sequences are in accord with the phonotactic rules of German. The metronome frequency used to control the speech rate was varied in steps of 5 from 90 to 110 beats per minute. In this way we got five items in any of the following three sets:

- (1) /'e:te 'a:te 'i:te 'o:te 'u:te/
- (2) /'pe:te 'pa:te 'pi:te 'po:te 'pu:te/
- (3) /'ple:te 'pla:te 'pli:te 'plo:te 'plu:te/

Segment, syllable and foot durations were measured on sonographic displays (see below). For the perception experiment we

removed the first and the last foot from these fifteen utterances. The utterances were combined in pairs in the following way (forming different subtests): The item with mean speech rate of set 1 first, i.e. /'a:te 'i:te 'o:te/ at a rate of 100 feet per minute, followed by one item of set 2, or one item of set 1 followed by /'pa:te 'pi:te 'po:te/ at the mean rate of 100, and both combinations in reversed order, i.e. /'pa:te 'pi:te 'po:te/ in first position and /'a:te 'i:te 'o:te/ second. In the same way set 2 was combined with set 3 and set 1 with set 3, resulting in 54 stimulus pairs. The stimuli were presented to a group of 19 subjects in an AX-format (same/different rate of speech) six times each in randomized order.

RESULTS

Acoustical Analysis: Durational Measurements

The measurements of the relevant parts of the utterances used in the German perception experiments were made using broad band sonographic displays. The following segments were measured independently:

prestressed consonants from the beginning of the fricative noise till voicing onset of the vowel (/f/, /p/) or till the end of the /l/-obstruction marked by a clear increase of energy in the higher formants; stressed vowels till the /t/-occlusion; stressed syllables from /f/-onset till offset of voicing of the vowel; unstressed syllables till onset of /f/-fricative; and single feet from one onset of /f/-fricative till the next.

Foot duration exhibits no difference between the different segmental compositions. It is always slightly longer than demanded by the presented metronome rate (mean 3.1%), i.e. the subject always is a little bit slower than the presented metronome pattern. Foot compression of

18.6% between the extreme metronome rates is in the range of the computed value (i.e. the relation 110/90: 18.2%). As expected, the different parts of the foot contribute differently to this overall compression: with 22.2% it is stronger in the unstressed than in the stressed syllable (16.5%), and in the stressed vowels with 18.8% it is stronger than in the prestressed consonants (13.5%). Computed over all rates of speech, single two-factorial analyses of variance showed that the duration of the stressed syllable is significantly determined by the vowel ($F(2,8) = 6.31$; $p < .05$) and the initial consonants ($F(2,8) = 38.22$; $p < .001$): syllables containing /a:/ and /o:/ (with durations of 393.7 and 392.3 msec) are longer than those containing /i:/ (375.3 msec) and /pl_/-syllables (408. msec) are significantly longer than /p_/-syllables (394.3 msec) which in turn are significantly longer than /_/-syllables (359. msec). These intrinsic differences at the foot level are compensated for in the given material by the reversed effects in the unstressed syllables (effect of preceding vowel: $F(2,8) = 11.54$; $p < .01$; of preceding consonants: $F(2,8) = 71.3$; $p < .001$): here the /te/ following a syllable containing /i:/ is longer (247.3 msec) than those containing /o:/ and /a:/ (230.3 and 229.7 msec), and /te/ following /p_/_ (227.3 msec) is longer than /te/ following /pl_/_ (219. msec).

At the segmental level one can see that even vowel duration is a complex function of the vowel itself ($F(2,8) = 44.39$; $p < .001$), the prestressed consonants ($F(2,8) = 4.42$; $p \text{ ca. } .05$), and an interaction of both factors ($F(4,16) = 3.74$; $p < .05$). In general /a:/ (213.3 msec) is longer than /o:/ (196.7 msec; with the exception of /a:/ and /o:/ following /f/, where both are not significantly different), which in turn is longer than /i:/ (171.7 msec).

Perception Experiments

The results of the different subtests (see above) are shown in Fig. 1-3. For further analysis the median of the 'same'-response distribution was computed for every subject in all subtests. A two-factorial analysis of variance showed a significant effect of set combination ($F(2,198) = 13.48$; $p < .001$): the results suggest that /'pa:te .../ would have to be produced at a rate of 98.95 to be perceived as fast as /'a:te .../ at a rate of 100 and, not significantly differing from this effect, that /'pla:te .../ would have to be produced at a rate of 98.74 to be perceived as fast as /'pa:te .../ at a rate of 100, but

/'pla:te .../ compared with /'a:te .../ at a rate of 100 would have to be produced at the significantly slower rate of 97.5 to be perceived as equally fast. All computed rates are different from 100, the rate they are compared with ($p < .01$). An effect of the order of presentation within stimulus pairs, clearly visible in a pretest with real sentences is not to be seen in the results of the analysis of variance.

We replicated part of the experiment with American-English material and subjects ($N = 9$). The initial /f/ in the material was replaced by /s/. Only the combinations of set 1 with set 2 and set 1 with set 3 were tested in the same way as before. The results are shown in Fig. 4 and 5. We can see an effect of order of presentation in Fig. 5: stimulus pairs with the simpler /'sa:te .../-sequence in first position (open columns) result in very rare 'same'-responses that never reach 50%, whereas in the reversed order (filled columns) we have more 'same'-responses, exceeding 50% when the simpler sequence is maximally faster than the complex one, but both response functions are cut off at this stimulus pair (to be seen as well in Fig. 4 and to a less degree also in the German results of Fig. 3). This order effect means that the second part of the stimulus-pair is heard as faster than the identical one in first position probably due to a normally given slowing down at the end of utterances. Because the 'same'-response function is cut off at the 90-100, 110-100 pairs the order effect cannot become visible in the results of the analyses of variance based on the median measurements: these do not represent the actual point of perception of equal speech rates.

Parallel to the German results the analysis of variance shows a clear effect of set combination on the median of 'same'-responses ($F(1,68) = 9.23$; $p < .01$): /'spa:te.../ would have to be produced at a rate of 98.24 to be perceived as fast as /'sa:te .../ at a rate of 100, whereas /'spla:te .../ would have to be produced at a rate of 96.12 to sound as fast as /'sa:te .../ at a rate of 100 (both computed rates differing from 100; $p < .01$). Because of the reasons mentioned above the analysis of variance again does not show a significant effect of order of presentation.

It should be mentioned that it is not possible to correlate the data of the perception experiments with the measurements of acoustical segment durations since the median-based results of the

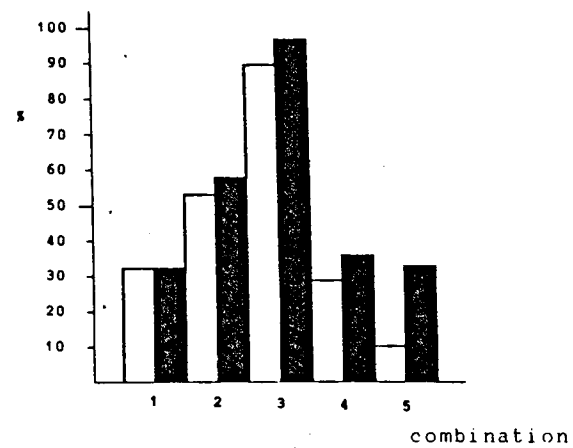


Fig. 1: /f.../ - /p.../ (open columns) and /p.../ - /f.../ pairs (filled columns)

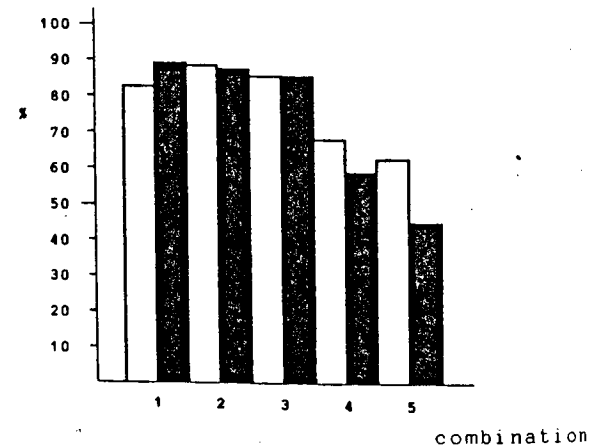


Fig. 4: /s.../ - /sp.../ (open columns) and /sp.../ - /s.../ pairs (filled columns)

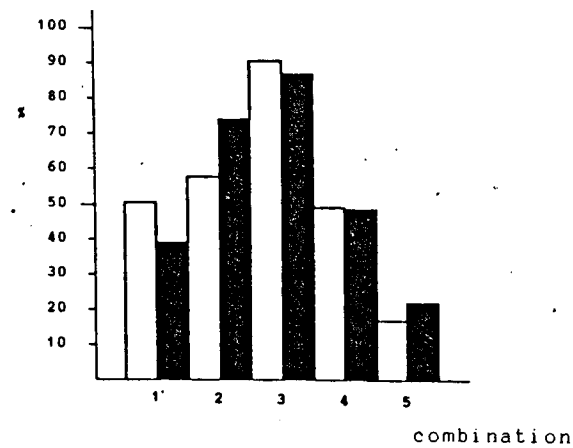


Fig. 2: /p.../ - /pl.../ (open columns) and /pl.../ - /p.../ pairs (filled columns)

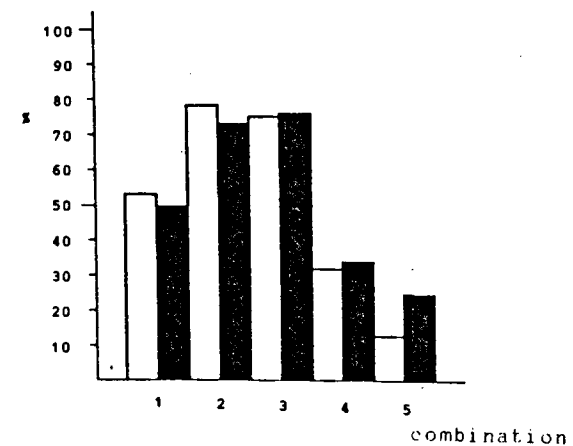


Fig. 3: /f.../ - /pl.../ (open columns) and /pl.../ - /f.../ pairs (filled columns)

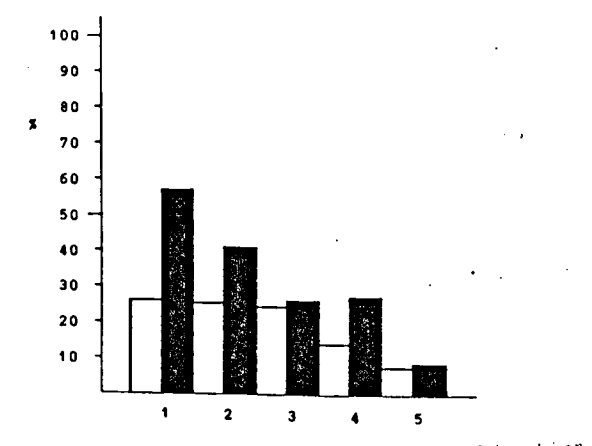


Fig. 5: /s.../ - /spl.../ (open columns) and /spl.../ - /s.../ pairs (filled columns)

perception experiments do not represent the actual measure of perceived equality of speech rate.

DISCUSSION

The durational measurements have shown that there was a good approximation of the metronome rate at the level of foot duration. But this durational compression is seen to work differently in different parts of the foot. Intrinsic durational differences of the stressed syllables for example are compensated for by the durational behaviour of the unstressed syllable. Although at the syllable level there are durational differences due to the complexity of the initial consonance up to 50 msec, the complex utterances are perceived as being uttered at a faster rate of speech. Clearly this judgement of the hearer must be based on a measure of articulatory movements per unit time.

REFERENCES

- [1] Hoequist, C. 1983, Parameters of speech rate perception. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)* 20, 99-138.
- [2] Kohler, K.J. 1986, Invariance and variability in speech timing: from utterance to segment in German. In: Perkell, J.S. & Klatt, D.H. (ed.), *Invariance and Variability in Speech Processes* (Hillsdale, N.J.), 268-299.