

THE TYPOLOGY OF SPEECH SEGMENT UNITS

R.K. Potapova

Moscow State Linguistic University, USSR

ABSTRACT

The study of speech communication presupposes a preliminary segmentation involving the identification of cues segments and their characteristics. The task of the present investigation consisted in the determination and comparative description of different speech segments correlated in duration in spoken utterances for German, English, Swedish and Danish.

1. INTRODUCTION

The organization of spoken connected speech implies the selective extraction of linguistic objects which is impossible without segmentation. There are three principal difficulties arising in the process of solving a task of speech recognition both in the context of natural speech communication and in the application of automatic recognition devices: optimum segmentation of an object which is very difficult for the units to be recognized have no clear-cut boundaries; accounting for the variability factor of object characteristics (for instance, variability of articulation and acoustic characteristics of a speaker); identification of a set of key characteristics. The orientation to -

wards key segments enables us to differentiate between micro-, medi- and macrosegmentation. In microsegmentation key segment we include audial and acoustic intrasound/intersound transient processes, occlusions, frictions, explosions, subsonic and sound segments, syllabic segments. The following can be cited as segmentation universals at the audial level: segmentation at the neuron level with the orientation towards changes on the domain of F_0 ; reaction on the maximum values of spectrum energy and changes in timing of speech signal energy. Certain acoustical characteristics of the microsegmentation can also be defined as universals: presence/absence of F_0 ; instantaneous change of F_0 at the transition from a consonant to a vowel and from a vowel to a consonant; presence/absence of spectrum noise; noise localisation at the frequency scale; noise intensity; noise duration; presence of low and/or high frequency spectrum energy. The objects of macrosegmentation are phrases, sentences and fragments of spoken text. Between the objects of micro- and macrosegmentation is the phonetic word (the accentual group with proclitic and enclitic syl-

lables), which makes it possible to separate missegmentation, where we are guided by the following characteristics: specific realization of prosodic and spectral characteristics of juncture sounds; time correlation between segments within phonetic word; integral intensity of phonetic word; qualitative and quantitative characteristics of stressed/unstressed vowels within phonetic word. In implementing the macrosegmentation of speech prosodic characteristics are the principal ones as well as an account for syntactic and semantic information.

2. PROCEDURE

The investigation was based on the German, English, Swedish and Danish material. We have recorded 20 speakers male and female for each language. The experimental corpus contained a set of interlingual identical words which were included into a set of sentences having identical rhythmic and syntactic structures. All acoustical characteristics were extracted at the first stage by intonograph and at the second stage by MICRO SPEECH LAB-ver. 3. One aim of our investigation is to systematically describe and to define realizations of duration of different speech segments in connected text. To discover correlation for duration data of different speech segments the correlation coefficient $\bar{\rho}$ was determined:

$$\bar{\rho} = \frac{\sum_{i=1}^n x_i y_i - \bar{x} \bar{y}}{\sigma_x \sigma_y}$$

The H_0 -hypothesis of the independence of relation of data was tested by means of t-criterion with $p=5\%$. It was necessary to determine: a) the nature of the time

correlation between the adjoining subsonic segments within syllable of the following types: occlusion (separately for voiced/unvoiced segments), explosion, friction, aspiration, transition; b) the nature of the time correlation between adjoining sound segments within phonetic word; c) the nature of the time correlation between the syllables within phonetic word; d) the nature of the correlation in duration between unstressed and stressed vowels within sentence.

3. DISCUSSION

In the course of the study of speech behaviour it was discovered that the temporal and structural organization of speech movements calls for a complicated speech programme in the brain. It should be noted in this connection that at present the following problems are being studied in the domain of the temporal organization of spoken utterance: -the determination of synthesis algorithms of rhythmic pattern of the utterance (the problem which is often referred to as preprogramming of the temporal organization of speech signal being part of a more general problem of synthesis (generation) of prosodic patterns of an utterance; -the determination of speech signals used in the information exchange between linguistic and physiological levels of speech analysis and synthesis; -the determination of the rules for the conversion of transformation of the rhythmic patterns of an utterance into real time intervals between articulatory positions and movements [1]. As a result on the data pro-

crossing it was demonstrated that the level of subsegment allowed us to establish a specific temporal correlation between such segments as, for instance, a voiceless occlusion and following frication. In Swedish a tendency was observed for establishing a statistical relevant correlation between the duration of a voiceless occlusion of a tense occlusive consonant and the duration of the following frication segment. In Danish a regular correlation was traced between the duration of the frication segment of a tense occlusive consonant and the duration of the following and the preceding vowels [2]. It was demonstrated that the consonant and vowel segments are characterized by different degrees of a sufficiently reliable temporal correlation. Thus, for instance, it was proved that in German and English exists a negative temporal correlation between sound segments in the VC - sequence. The English language was characterized with the negative temporal correlation between the vowel and the following consonant in the syllable and phonetic word irrespective of its position in sentence. In German the correlation of the same type was revealed for the sequence "vowel-consonant", but it was mainly observed in the final position of sentence. It should be noted that this data were gotten for two-syllable words with a short stressed vowel. For the long stressed vowels in German there exists temporal correlation between a long vowel and a preceding consonant. No such correlation was observed between a long vowel and a following

consonant. The temporal correlation established for the "vowel-consonant" sequence underwent no substantial change dependent on position, which show a relative stability of the temporal relations between the sound segments which form closed syllables in the languages. The data obtained indicate that the pattern of temporal compensation at the microsegmentation level is not substantially modified even in those cases, when the sound - and subsegment sequence becomes part of a rhythmic structure of a higher linguistic order. This rule is typical for connected speech in German and English. In Swedish the situation is somewhat different, and the temporal correlation may be registered for sequences of sound segments "consonant-vowel", as well as for sound segments "vowel-consonant". This type of temporal correlation was observed not only in different words but also within the same word. In Danish the temporal correlation should be noted between sound segments only in the sequences "vowel-consonant". It follows from all that was said above

that the temporal relation at the level of microsegmentation within a syllable in connected speech is looser in some languages and stronger in other languages. This means, that the involvement of sound segments into a sequence of the speech continuum may substantially modify the type of the temporal relation between adjoining microsegments, or may have no significant impact leading to a temporal redistribution. In the former

case, the temporal relation between the constituents of a syllable is sufficiently mobile, which brings about new types of temporal correlation, while in the latter case, the temporal relation between the constituents is sufficiently strong, which preserves the relative integrity of the syllabic structure in the connected speech. This conclusion appears rather well-founded, - which is an evidence of natural flexibility of the syllable as a fundamental material quantum, on which the whole speech "building" places. For the microsegmentation positive temporal correlation between syllables of different type within a phonetic word is characteristic of some languages. The temporal correlation was observed between CVC and CV-segments within phonetic word. No correlation was observed between other types of syllable combinations: GV -(C)C.V within the same phonetic words in utterances. The comparison of duration data for vowel - sequences in sentences in case of macrosegmentation revealed a positive temporal correlation between them. The data obtained suggest that the duration of units of micro, - medi- and macrosegmentation in connected speech may be realized with different degrees of regularity and hierarchical character. It may be concluded that the architectonics of a speech utterance does not merely amount to a simple sum of duration data of a set of micro, - medi, - macrosegments, but emerges instead as a more complex structure, comprising

some relatively autonomous units, the temporal organization of which is predetermined by their own micro, - medi- and macrosegmental properties, as well as by the corresponding prosodic properties of the whole speech structure in general [3].

4. CONCLUSION

The variability of the prosodic organization of a utterance is brought about by the interaction of the following factors: the physiological one as predetermined by the constitution of human speech organs; the physiologically-linguistic one as predetermined by the laws of coarticulation in accordance with the features of the pronunciation basis of a given language; the linguistic one as predetermined by the phonemic, morphemic and syntactic rules of a language; the psycholinguistic one as predetermined by the communication act as a whole. The above mentioned factors are overlapping in the process of temporal programming and all types of segmentation of a spoken utterance.

5. REFERENCES

- [1] ČISTOVIČ, L. (1972), "Problemy issledovanija vremenoj organizatsii reči" Leningrad.
- [2] POTAPOVA, R. (1986), "Slogovaja fonetika germanских jazykov", Moscow, Vysshaja škola.
- [3] POTAPOVA, R. (1990), "Strukturno-tipologičeskoje segmentovedenije zvučasnoj reči", Tipologija jazykov, Moscow, MGLU, 70-87.