THE INFLUENCE OF THE SYLLABLE BOUNDARY ON CONSONANT-CONSONANT REALIZATIONS

Thomas Portele
Institut für Kommunikationsforschung und Phonetik, Universität Bonn

ABSTRACT

This paper describes an investigation on the influence of the syllable boundary on consonant-consonant pairs. About 2000 pairs extracted from continuos speech were analyzed. In the acoustic domain it was found that the syllable boundary has some slight influence insofar, as sounds after a syllable boundary tend to be pronounced more clearly. In the perceptive domain we found a corresponding ability to place the syllable boundary. However, both findings are not very marked.

MOTIVATION

The experience with our syllable-based speech synthesis system HADIFIX [6] indicated that the syllable boundary does not serve as a coarticulation blocker; instead we found that our system often generated hypercorrect and over-articulated speech because effects happening a the syllable boundary are not taken into account by just concatenating demisyllables. In the course of the definition of a better inventory structure [7] the question arose whether the syllable boundary has any influence on the phonetic-acoustic domain. Of course, the nature and strength of this influence was what we searched for.

METHOD

Consonant-consonant pairs were placed into 48 specifically designed short texts. All possible combinations occured in different frequencies determinated by the structure of German; the pair /kl/ appeared more often than the pair /tl/. If not prohibited by German phonotactics, the position of the syllable boundary varied between initial (preceding a consonant pair as in "betragen" /bə.tra:gən/) and medial position (between the two consonants as in "mitreißend" /mɪt.raɪsənt/), and between medial and final position (following a consonant pair).The complete corpus contained more than 2000 pairs. It was read by a woman and by a man who were unaware of the study's aim. The pairs were extracted, segmented and grouped according to the way some features were realized. For example, phonologically voiced sounds were labelled *completely devoiced*, *partially devoiced* and *voiced*; stops were tagged *not spoken*, *not released*, *weakly released*, *released*, and *released and aspirated* [2]. Fricatives, nasals and liquids were labelled as well as the degree of assimilation to the other sound of a pair. These classifications were performed only by the investigator. However; a pilot study with 200 plosive-plosive combinations and an additional labeller as well as a re-labelling by the investigator confirmed the reliability of the classification (correlation coefficient ~ 0.9 for both inter-rater and intra-rater reliability). The temporal structure of the consonant-consonant pairs was also investigated. In some cases, spectral features were taken into account. All these results were statistically analyzed regarding the influence of the syllable boundary on the distribution of these parameters.

RESULTS

The results were obtained by comparing the distribution of the labels described above in dependence on the position of the syllable boundary. The analysis yielded experimental results about reduction, assimilation and other phenomena [3], however, only the results concerning differences caused by the syllable boundary are described here.

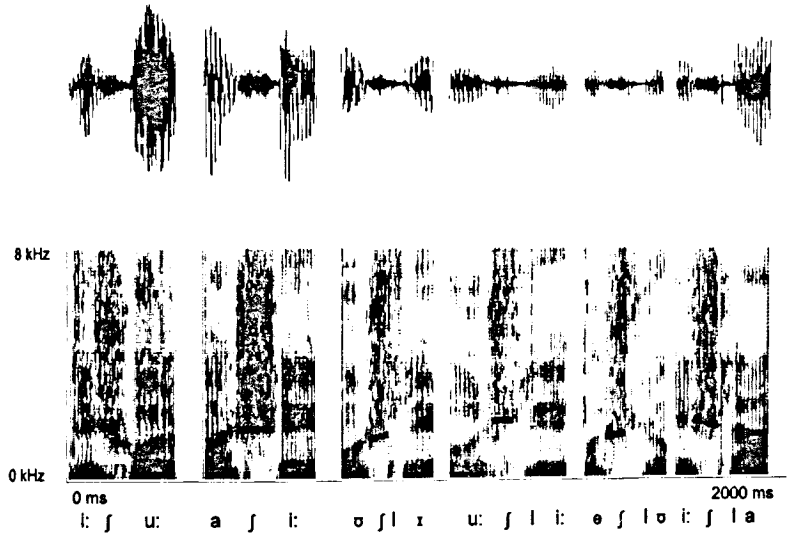

Figure 1. Context-dependent rounding of the /ʃ/ and its acoustic manifestation as a lowered formant-like structure (marked by the fat lines). A rounded /ʃ/ has a formant-like structure by 1450 Hz, an unrounded one by 1900 Hz. From left to right: /i:.ʃʊ/ (transition inside the fricative), /a.ʃi:/ (unrounded), /ʊʃ.lɪ/ (rounded), /u:.ʃli:/ (unrounded), /ɐ.ʃlʊ/ (rounded), and /ɪ.ʃla/ (unrounded). The dot denotes the phonological syllable boundary.

Plosives

The plosive-plosive combinations /pt./ and /kt./ (the dot denotes the position of the phonological syllable boundary) were compared to /p.t/ and /k.t/, respectively. It was found that the first plosive is articulated clearer in syllable final position (U-test, $p < 0.1$). A syllable final /p/ before /t/ shows longer closure duration than /p/ in /pt./.

For the plosive-fricative combinations /kv,pf,ps,tʃ/ significant effects were found only for /kv/, where a syllable initial /v/ is more often voiced or partially devoiced compared to /v/ after a syllable initial /k/ (U-test, $p < 0.025$). There is a tendency for plosives to be released more forcefully in syllable initial position; however, this phenomenon was not significant.

The combinations /.kn/ and /k.n/ differ regarding to the voice onset time of the /k/ which is longer in syllable initial position (U-test, $p < 0.05$).

In plosive-liquid combinations the liquid is less likely to be devoiced when in syllable initial position. This tendency is not significant in our data.

Fricatives

In the plosive-fricative combinations /ʃt/ and /ʃp/ the /t/ is aspirated in syllable initial position and unaspirated otherwise (U-test, $p < 0.05$). A syllable initial /ʃ/ is shorter (U-test, $p < 0.1$).

The position of the syllable boundary has no effect on the combinations /fs,ʃv,ʃm,ʃn,fr,fl/.

In /ʃ.l/ and /ʃ.r/ the /ʃ/ is longer than in /.ʃl/ and /.ʃr/, respectively (U-test, $p < 0.1$).

An interesting result is given in Figure 1. Displayed are different versions of /ʃ/ in rounded and unrounded contexts. One can easily see that the feature *rounded* is determined by the vowel that belongs to the same syllable as the /ʃ/. However, this result was not reproducable for other fricatives (but both speakers in our database produced the phenomenon for the /ʃ/).

Sonorants

In combinations with a sonorant as the first sound no significant influence of the syllable boundary could be established in our data.

DISCUSSION

The influence of the syllable boundary on consonant-consonant combinations can be found for some consonant pairs. However, it does not seem to be a strong one. No categorical differences between syllable initial and non-initial versions of suitable sounds could be observed (except for the rounded /ʃ/); sounds after a syllable boundary just tend to be pronounced more clearly. And even this might be an artifact of the test material when one assumes that the duration of a sound in a syllable or syllable part is inversely correlated to the number of sounds in the respective unit [5]. The duration of a /ʃ/ in a one-sound syllable onset is longer than in a two-sound onset. This might not be a specific influence of the position of the syllable boundary but merely a timing universal that is sometimes called *isochrony*. And, if a sound is longer, it can be pronounced more accurate [4]. It seems that the influence of the syllable boundary can be called quite negligible in German. Lip rounding, however, might be an exception, but, as long as the effect is not better reproduced (just for one sound in speech from two speakers), it might just be an oddity.

PERCEPTUAL TEST

The acoustic analysis of the data showed an influence of the syllable boundary that can be regarded largely as a timing effect. To assess whether human listeners are able to use this effect a small experiment was carried out.

Method

Ten consonant pairs were selected, namely /kt, kv, fl, kl, pf, ʃp, ʃt, ʃv, tʃ, ts/. Each pair appears in our data in two types, one with the syllable boundary between the consonants, and one with the boundary after (/kt/) or before (all other pairs) both consonants. Two versions from each type were chosen randomly. They were extracted together with the surrounding vowels. Each of these stimuli appeared twice in the test. Altogether, 80 pairs were judged by the subjects.

Each stimulus was played twice. The test was recorded on cassette and played to the subjects by earphone in a quit room. Eleven subjects participated; most of them were trained phoneticians.

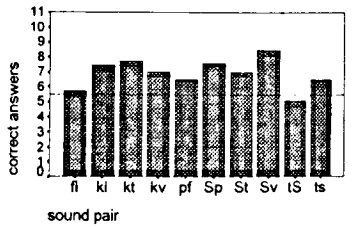

Figure 2. Results of the perceptual test displayed as correct answers for a stimulus. The horizontal line indicates the chance value.

Results

Figure 2 gives the results for each sound pair. It is easy to see that the results are better than chance value but that the correct answer rate is not overwhelmingly high. Among the subjects there are clear differences (Figure 3). Most subjects described the test as difficult; they said that they tried to guess the words according to the surrounding vowels and deduce the position of the syllable boundary from their guess applying their linguistic knowledge.

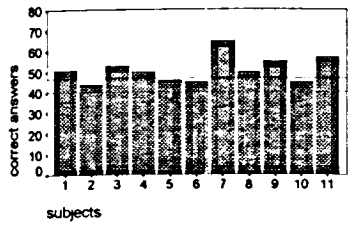

Figure 3. Results of the perceptual test displayed as numbers of correct answers from each subject. The lower horizontal line indicates the chance value, the upper horizontal line denotes the value, where the difference to the chance value becomes significant (χ^2-test, $p < 0.05$).

Discussion

Relaying on linguistic instead of phonetic hints showed to be a quite successful strategy. Therefore, the exploitable phonetic information must have been negligible, a fact that is also demonstrated by the poor recognition rates of some subjects.

GENERAL DISCUSSION

The rather obscure status of the syllable in the phonetic science [3,8] seems to be justified by the rather obscure results described above. A clear influence of the position of the syllable boundary on the properties of consonants could be established neither in the acoustic nor in the perceptual domain. The perceptual data might be consistent with some kind of low level perception of syllables as a whole that can not be applied consciously by the subjects under the artificial test conditions. But in combination with the acoustic data it is obvious that the results confirm the dubious state of the syllable in speech. However, only segmental effects were under investigation, and the syllable is an important prosodic unit [1,5]. It is difficult to distiguish between temporal and spectral domains because *target undershoot* phenomena and other reductions are a direct byproduct of a shorter sound [4].

The application of these results to concatenative speech synthesis yields to the conclusion that the syllable boundary can be neglected when choosing the units for concatenation. However, a sophisticated temporal control is necessary that does not only adjust the segmental durations according to the position in the syllable, but that supports a nonlinear timing control; for instance, aspiration is more likely to be shortened than closure duration.

ACKNOWLEDGEMENT

I thank Florian Höfer for assistance with the material, and Gerit Sonntag for carrying out the perceptual test. The research was funded by the *Deutsche Forschungsgemeinschaft.*

REFERENCES

[1] Campbell, W.N.; Isard, S.D. (1991) "Segment durations in a syllable frame." Journal of Phonetics **19**, 37-47

[2] Henderson, J.B.; Repp, B.H. (1981) "Is a stop consonant released when followed by another stop consonant?" Phonetica **39**, 71-82

[3] Kohler, K.J. (1977) *Einführung in die Phonetik des Deutschen.* Berlin: Erich Schmidt

[4] Lindblom, B. (1963) "Spectrographic study of vowel reduction." J. Acoust. Soc. Am. **35**, 1773-1781

[5] Meyer, H.; Portele, T.; Heuft, B. (1995) "Ein Silbendauermodell für die Sprachsynthese." (to appear in: *Fortschritte der Akustik - DAGA'95).*

[6] Portele, T.; Steffan, B.; Preuß, R.; Sendlmeier, W.F.; Hess, W. (1992) "HADIFIX - a speech synthesis system for German." Proc. ICSLP'92, 1227-1230

[7] Portele, T.; Heuft, B.; Höfer, F.; Meyer, H.; Hess, W. (1994) "A new high quality speech synthesis system for German." Proc. CRIM/FORWISS-Workshop on Speech Research and Technology, Munich, 284-287

[8] Tillmann, H.-G. (1964) *Das phonetische Silbenproblem.* Diss., Universität Bonn