THE SYLLABLE AS A PERCEPTIVE CRITERION TO DISTINGUISH BETWEEN VOWELS AND 'GLIDES'

Doris Mücke*, Frank C. Stoffel*, and Bernd J. Kröger†
*Institute of Phonetics, University of Cologne, Germany, †Research Centre for General Linguistics, Berlin, Germany

ABSTRACT

This study deals with the syllable as a reliable unit to distinguish between 'glides' and vowels. Starting from a gestural framework implemented in an articulatory speech synthesizer we generated stimuli varying from syllabic to non-syllabic /i/-sounds in prevocalic positions of German. Avoiding qualitative effects we reduced the corresponding vocalic gestures in their magnitude by manipulating the gestural eigenperiod which controls the velocity of the gesture-executing articulator towards its spatial target. The transition of syllabic (vowel) to non-syllabic ('glide') function were tested by two groups of listeners: 150 native speakers of German and 69 speakers of German as a second language.

1. INTRODUCTION

Syllables are considered the basic units of speech. While native speakers seem to have an intuitive knowledge about the number of syllables in a given utterance in their own language, phonetic definitions and phonological concepts of the syllable are problematic; thus it is impossible to find clear syllable boundaries in spoken language although they ought to result from segmental analysis [1].

However, in speech perception the syllable seems to be a reliable unit. We used it to distinguish between 'glides' and their vowel counterparts. Phonetically, the quality of a 'glide' is that of the corresponding vowel, but they differ in duration [2]. In this study the phenomenon of the transition of a syllabic (vowel) to a non-syllabic ('glide') function of these sounds is discussed on the basis of German palatal 'glides' by means of an articulatory speech synthesizer and perception tests.

1.1. Definition

The term 'glide' might be misleading since it also refers to transitional sounds (on-glide, off-glide) described e.g. by Sievers [3]. In this study 'glide' is synonym for semivowel, which is defined as a vowel-like sound/segment which functions as a consonant [4].

2. THE GESTURAL FRAMEWORK

2.1. Articulatory Phonology

In articulatory phonology (Browman and Goldstein) syllables as hierarchical units are modelled by the mechanism of phasing of individual gestures, i.e. no syllable boundaries need to be determined [5]. Gestures are defined as basic units of phonological contrast and as abstract characterizations of articulatory events. As primitive phonological units gestures specify discrete categories, e.g. bilabial constriction or lip protrusion, as units of events they determine articulatory movements, each with an intrinsic time of duration. Every articulatory movement is aimed at a target containing a linguistically relevant constriction. Gestures may overlap in time. They don't correspond to known segments or features. The mere presence or absence of a given gesture may e.g. change a sound distinctively (Browman and Goldstein):

(a) <u>add</u> and <u>bad</u> differ in the bilabial constriction of the lips
(b) <u>pad</u> and <u>pan</u> differ in the aperture of the velum

Most of the phonological structure can be represented by gestural constellations (phasing); in our case the transition from syllabic to non-syllabic of prevocalic /i/-sounds in German is modelled without adding or removing any gestures.

2.2. A Quantitative Speech Production Model

The gestural framework was implemented in an articulatory speech synthesizer by Kröger [6]. It is a quantitative speech production model generating fluent speech without explicit segmental rules. The simulation of increasing speaking rate e.g. is possible by gestural phasing leading to a huge number of different reduced forms; the reduction degree can be read from the phase values directly. The articulatory speech synthesizer comprises two components: the dynamic model generating the articulatory movements and the articulatory-acoustic model forming the midsagittal shapes of the vocal tract and generating the audio signal. The dynamic model provides control parameter time functions including phonatory control parameters like glottal aperture, pulmonary pressure and vocal fold tension as well as articulatory parameters controlling the articulators lips, tongue and velum.

In a quantitative speech production model the articulatory movements are specified by gestural scores. There are three types of gestures: vocalic, consonantal and opening gestures. Consonantal gestures are associated with start phases and end phases of the vocalic gestures: They have a shorter intrinsic duration and are completely overlapped by the vocalic gestures.

2.3. Simulation of Increasing Speaking Rate

In fluent speech, the occurrence of reduced forms increases with the speaking rate. Two gradient modifications to gestural structure were described by Browman and Goldstein: (a) increase in overlap and (b) decrease in gesture magnitude. No segments are added or removed. In an articulatory speech synthesis system it is possible to generate a variety of quantitative alteration processes with different degrees of reduction by gestural phasing. Two strategies are available to simulate increasing speaking rate [7]:

(a) Decrease in gesture magnitude
Strategy: Decreasing articulatory velocity by decreasing the parameter gestural eigenperiod.
Result: Manipulation of sound duration without qualitative effects
(b) Increase in overlap of two gestures
Strategy: Increasing of gestural overlapping in time by decreasing the parameter release phase.
Result: Manipulation of sound duration with qualitative effects.

In a gestural production model phase values don't give the absolute time, but the degree of realisation of each gesture. Simulation of increasing speaking rate involves the quantitative parameters 'eigenperiod' and 'release phase'.

Eigenperiod determines the associated time interval within which an intended given gestural target should be approximated. With a decrease of eigenperiod the articulator moves faster towards its spatial target and thus decreases sound duration without influencing the pattern of articulation. The release phase on the other hand determines the phase value at which the gestural activation ends. With a low release phase value the degree of realisation of the active gesture is low, and the temporal overlap of gestures increases; as a result the pattern of articulation is influenced qualitatively, up to segmental modification.

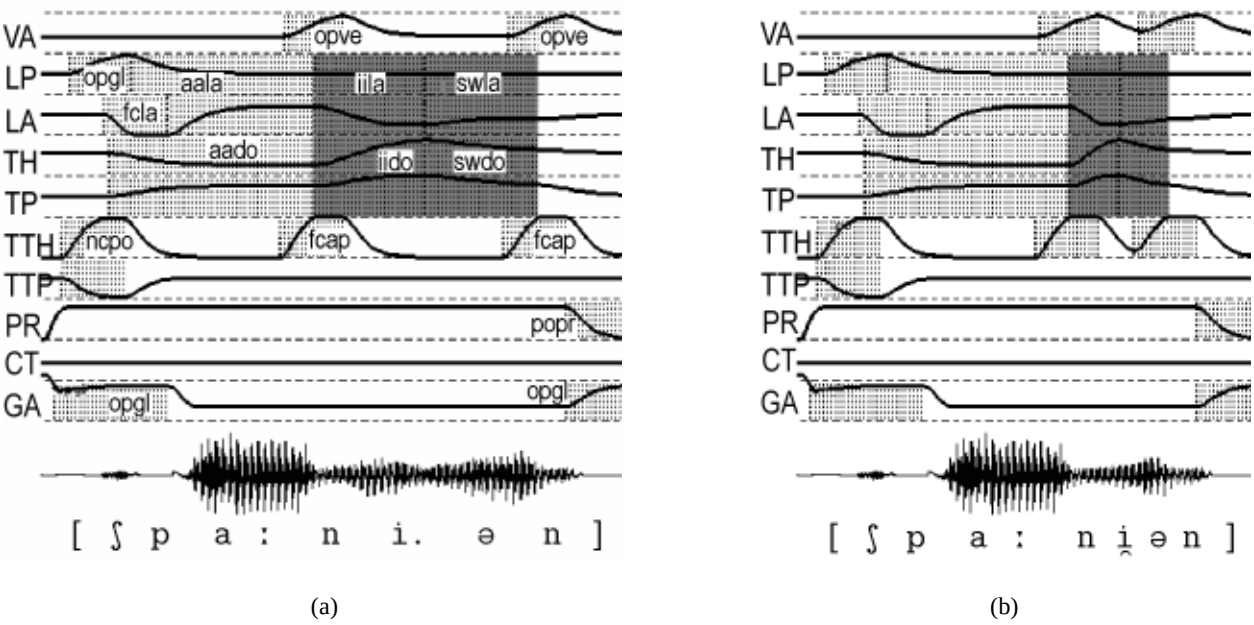
3. CASE STUDIES: SYLLABIC VS. NON-SYLLABIC

We generated stimuli for German <Spanien> and <Dahlien> varying from syllabic to non-syllabic /i/-sounds in prevocalic position without qualitative effects. Therefore we used the first strategy (manipulating articulatory velocity without qualitative effects) and changed the parameter eigenperiod. Perception tests show that at a critical degree of gestural extension, the sound looses its syllabic function, and the number of syllables decreases by one.

3.1. Stimuli

Eight stimuli were generated for each word (<Spanien> and <Dahlien>) with eigenperiods varying between 75 degrees and 250 degrees at constant intervals of 25 degrees. Figure 1 shows control parameter time functions and oscillogram for the synthetic speech signal of <Spanien>:

The control parameter time functions illustrate the movement of the articulators (black curves). Each column stands for an associated articulator: velic aperture (VA), lip protrusion (LP), lip aperture (LA), tongue height (TH), tongue position (TP), tongue tip height (TTH), tongue tip position (TTP), pulmonary pressure (PR), cord tension (CT) and glottal aperture (GA). Each gesture is active for a distinct time interval; the gestural activation intervals are presented in shaded areas. The gesture name stands for the respective gesture target: labial full-closing gesture (fcla), apical full-closing gesture (fcap), postalveolar near-closing gesture (ncpo), the opening gestures of velum (opve) and glottis (opgl) plus the dorsal vocalic gestures to realize /a/ (aado), /i/ (iido), /ə/ (swdo) and the labial vocalic gestures /a/ (aala), /i/ (iila), /ə/ (swla). If a gesture is active, the associated articulator (black curves) executes a goal-directed movement towards a spatial target, which is never actually



(a) (b)

Figure 1. Control parameter time functions and oscillograms of <Spanien>.
(a) The eigenperiod value of the vocalic schwa- and /i/-gestures is set to 225 degrees which lengthens the synthetic signal of <Spanien> so that it is judged as a stimulus with three syllables.
(b) The eigenperiod value of the vocalic schwa- and /i/-gestures is set to 100 degrees which shortens the synthetic signal of <Spanien> so that it is judged as a stimulus with two syllables.

reached but approximated. Having approximated its respective target the articulator moves either towards a new target or towards a neutral schwa-like position.

The modification of eigenperiod leads to a modification of articulator velocity. A gesture with a low eigenperiod approaches its target faster than a gesture with higher eigenperiod. The synthesizer presets the average eigenperiod to 160 degrees for vocalic and to 80 degrees for consonantal gestures. For the test stimuli we manipulated only the eigenperiod of the schwa- and /i/-gestures.

The gestural activation intervals determine indirectly the duration of sounds of the synthetic signals. The gestural activation intervals of the schwa-gestures are considerably shorter than the /i/-gestures, since schwa only occurs in unstressed syllables as a short vowel in German.

3.2. Perception Tests

Perception tests were performed with two groups of listeners: native speakers of German and speakers of German as a second language. The synthetic stimuli of <Spanien> and <Dahlien> were judged in terms of 'number of syllables' per stimulus.

3.2.1. Procedure

Each stimulus (eight <Dahlien> and eight <Spanien> with varied eigenperiods) was presented three times in order to counteract possible chance decisions. The total of 48 synthetic signals was played back at intervals of three seconds in mixed order. The time interval was kept short to provoke a spontaneous reaction of the listeners.

The Stimuli were presented in an identification test with two answer categories. The question was: Do you hear *two* or *three* syllables in the following synthetic stimuli? Listeners were forced to choose between 'two syllables' or 'three syllables' for each test item [8].

Students and colleagues of the Institutes of Phonetics and German Philology (linguistic department) of the University of Cologne took part in the test. The stimuli were played back with a tape recorder to groups of listeners who put down their answers on prepared lists. They were also asked to give their age, sex, mother tongue (possibly dialects) and places of residence. The answers of listeners with mother tongues other than German were analyzed separately; incomplete answer sheets were excluded.

3.3. Results

150 test papers by native speakers of German were included in the first analysis. Each stimulus was judged 450 times (150 listeners x 3 repetitions). Figures 2a-b show the distribution of the two complementary answer categories.

The perceived number of syllables clearly depends on the

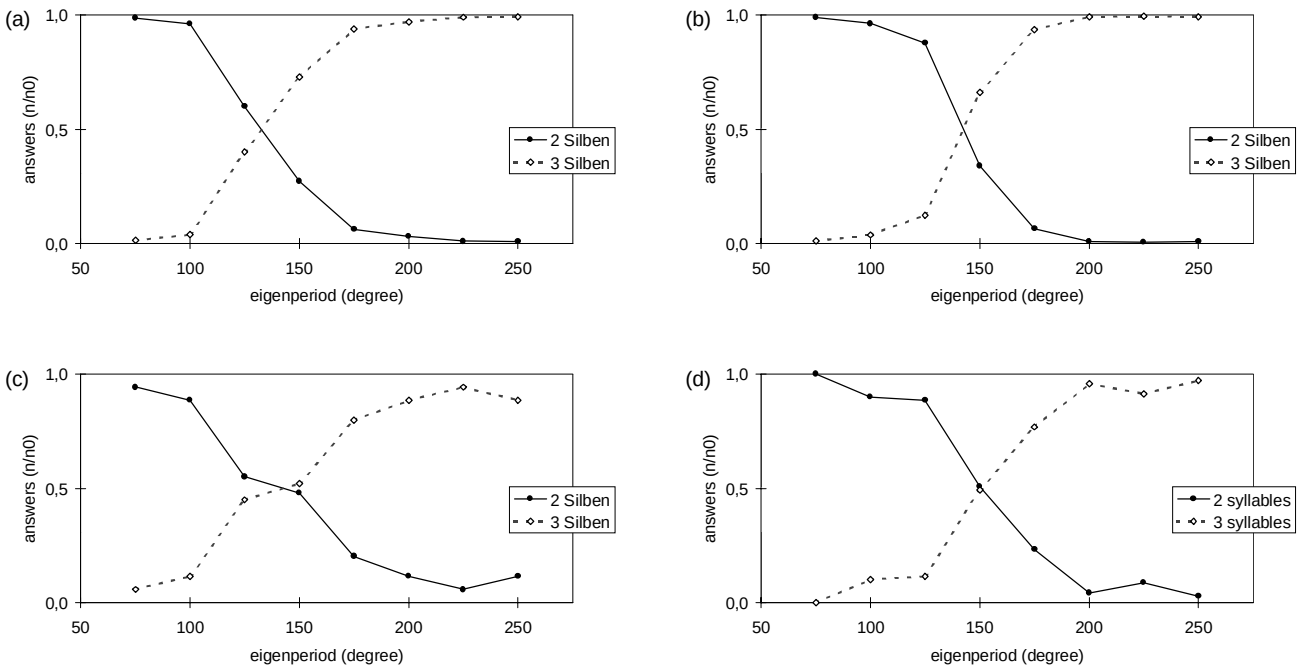


Figure 2. Listeners' judgments on the number of syllables ('two syllables', 'three syllables').
(a-b) Answers of 150 listeners with first language German. Perceived categories 'two syllables' and 'three syllables' switch at eigenperiod values between 125 and 150 degrees for <Spanien> (a) and 150 degrees for <Dahlien> (b).
(c-d) Answers of 69 listeners with German as second language. Perceived categories 'two syllables' and 'three syllables' switch at eigenperiod value of 150 degrees for <Spanien> (c) and <Dahlien> (d).

eigenperiod values. There is a relatively narrow range of chance decisions round 150 degrees; above that level, /i/-variations received syllable function, i. e. the number of perceived syllables increased by one. The increase (or decrease) of the perceived number of syllables seems to follow categories which, in the case of the prevocalic /i/-sound, depend on the relative sound duration.

For comparison, 69 test papers by listeners with mother tongues other than German were also analyzed. Among their mother languages were Russian, Swedish, Turkish, Hungarian, Korean, Japanese. Each stimulus was judged 207 times (69 listeners x 3 repetitions). Figures 2c-d show the distribution of the two complementary answer categories.

Again, the perceived number of syllables depends on eigenperiod values, i.e. on relative sound duration. The data bases of the two analyses (150 native vs. 69 non-native speakers of German) are too different in size to allow more detailed interpretation.

However, the non-native speakers of German also seem to have an intuitive knowledge about syllables in German. They may have acquired this knowledge in the course of their studies of German. On the other hand, their judgements might imply that the syllable is a universal unit shared by all languages as a fundamental concept. Further tests with listeners without any knowledge of German may help to decide that question.

4. CONCLUSION

This study shows that 'glides' are short non-syllabic vowels which differ from their syllabic counterparts *only* in duration. The syllable is a reliable criterion to distinguish between 'glides' and vowels, as they differ perceptively in their syllabic or non-syllabic function.

Starting from a gestural framework implemented in a quantitative speech production model we generated stimuli varying from syllabic to non-syllabic /i/-sounds in prevocalic position in German without affecting the sound quality. At a critical degree of gestural extension, the sound looses its syllabic function, and the number of syllables decreases by one.

Perception tests were performed with two groups of listeners: native speakers of German and speakers of German as a second language. Both groups agreed in their judgements on the number of stimuli syllables. The 'glides' and their vocalic counterparts were distinguished by their duration; in our case, the critical value (gestural eigenperiod of 150 degrees) held for both groups. The syllabic vs. non-syllabic function of vowels vs. corresponding 'glides' was valid for all listeners irrespective of their first language.

The study will be extended in two directions:

1. Synthesis: Stimuli will be created by decreasing the release phase; the resulting increase in overlap among gestures has qualitative effects: the prevocalic /i/-sounds are reduced towards schwa. Do vowel quality changes influence the auditory perception of syllable numbers?
2. Tests: Listeners without any knowledge of German are to be included in the test. Is listeners' judgement on the syllabic vs. non-syllabic function of vowels vs. 'glides' independent of their first language?

ACKNOWLEDGMENTS

Our thanks to Georg Heike and Heinz Vater for discussion and to Christine Riek for help in manuscript preparation.

REFERENCES

[1] Heike, G., 1992. Zur Phonetik der Silbe. In : Eisenberg P., Ramers K.H. and Vater, H. (eds), Silbenphonologie des Deutschen. Tübingen: Narr (=Studien zur deutschen Grammatik 42), 1-44.

[2] Mücke, D., 1998. Halbvokale im traditionellen Untersuchungsbereich der akustischen Phonetik. In: Noack, C. and H. Spieckermann (eds.). Studien zur akustischen Phonetik. Aachen: Shaker 1999 (to appear).

[3] Sievers, E., 1876. Grundzüge der Lautphysiologie. Zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen. Leipzig: Breitkopf und Härtel.

[4] Ladefoged, P. / Maddieson, I., 1996. The Sounds of the World´s Languages. Oxford: Blackwell Publishers.

[5] Browman, C.P. and L. Goldstein, 1992. Articulatory Phonology: An Overview. In: Phonetica 49: 155-180.

[6] Kröger, B.J., 1993. A Gestural Production Model and Its Application to Reduction in German. In: Phonetica 50: 213-233.

[7] Kröger, B.J., 1996. Zur phonetischen Realisierung von Sprechtempoänderungen unter Einbeziehung von artikulatorischer Reorganisation: Artikulatorische und perzeptive Untersuchungen. In: Gibbon, D. (ed.). Natural Language Processing and Speech Technology. Berlin: Mouton de Gruyter, 171-185.

[8] Greisbach, R., 1996. Experimentelle Testmethodik der linguistischen Phonetik - am Beispiel von segmentalen Grenzphänomenen im Deutschen. Institute of Phonetics, University of Cologne: Postdoctoral thesis.