ISCA Archive
http://www.isca-speech.org/archive

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

# INTRINSIC PROSODIC VALUES AND SEGMENTAL CONTEXT

Barbara Heuft & Thomas Portele
e-mail: bhe@asl2.ikp.uni-bonn.de
Institut für Kommunikationsforschung und Phonetik, Universität Bonn
Poppelsdorfer Allee 47; 53115 Bonn
GERMANY

## ABSTRACT

This investigation deals with two types of microprosodic variations: the intrinsic values for $F_0$ and duration in German and the influence of adjacent sounds on these values. The values are computed using an inventory for concatenative speech synthesis. The results of the measurement of intrinsic $F_0$ and duration for vowels correspond to those of other studies. The $F_0$ values for voiced consonants and durations for all consonants are studied as well. The influence of voicing, nasality and place of articulation of a preceding consonant on vowel $F_0$ and duration is described.

## 1. MOTIVATION

The present study is one step in our current work on prosody generation for speech synthesis. We need reliable information about intrinsic prosodic values. A lot of interesting work has been carried out in the past. Most studies focus on vowels only [1] [2] [3] or did not take into account the segmental context [1] [2]. Broader studies have been carried out for English and French [4], [7] but one cannot simply transpose results from one language to the other.

## 2. METHOD

The unit inventory for concatenative speech synthesis is at the same time our database to determine intrinsic prosodic values [5]. It contains all speech sounds in a large number of segmental contexts, spoken by a female speaker. Thus, it is possible to determine both the intrinsic values and the influence of the context at the same time. The demisyllables and diphones which made up the inventory were embedded in carrier phrases. All these units appeared in secondary-stressed position; a reduced pronunciation and a strong influence of macroprosodic effects is avoided. The inventory was recorded in an anechoic chamber; all carrier sentences were uttered with almost identical intonation. However, two types of macroprosodic context appeared: the end of a subordinate prosodic phrase (a noun phrase) and the middle of a phrase. The values for this contexts were investigated separately. It was possible to obtain some results about the interplay between macroprosody and intrinsic prosodic values.

Intrinsic $F_0$ was measured relative to a referenc which the mean values of the two syllables preceding the investigated segment. The duration data were extracted based on a segmentation done by hand.

For the statistic examination, we used non-parametric tests (Mann-Whitney U-test for two samples; Kruskal-Wallis one-way ANOVA for several samples.) A difference is called significant when p is smaller than 0.05.

## 3. RESULTS

### 3.1 Fundamental frequency

#### *3.1.1 Intrinsic values of vowels*

The results are consistent with the data from other studies ([1], [3]). Vowel intrinsic pitch depends on vowel height; open vowels have lower intrinsic pitch values than closed ones. This implies, at least for German, that long vowels have a higher pitch than their short counterparts. Among vowels with equal tongue height, back vowels show higher $F_0$ values than front vowels (see fig.1). However, there are two exceptions from the first rule: the long [a:] and [E:] in non-final position show a relatively high $F_0$ value. This might be a speaker dependent variation, but it could be a hint at a stronger influence of macroprosodic variation on open vowels. On the other hand, the values for [e:] are nearly identical in both contexts.

#### *3.1.2 The influence of the segmental context*

##### *3.1.2.1 Voicing*

Voicing of a preceding consonant decreases $F_0$ of the adjacent vowel, whereas voiceless consonants show the opposite effect. Devoiced consonants (i.e. phonologically voiced consonants produced without voicing after a voiceless sound) behave similar to voiceless consonants (see fig.2). Voicing of a following obstruent cannot be examined, because all plosives and fricatives in German are produced voiceless at the end of closed syllables.

##### *3.1.2.2 Manner and place of articulation*

The manner of articulation of preceding and following consonants does not have any influence on vowel $F_0$.
Sonorants cause a lower $F_0$, but this is due to the influence of the nasals in this group. [R] or [l] do not behave differently.
The place of articulation has a very small effect on vowel $F_0$. Labial consonants tend to elevate the $F_0$ values slightly, but this is significant only for the voiceless consonants. Alveolar or velar consonants do not show a consistent effect.

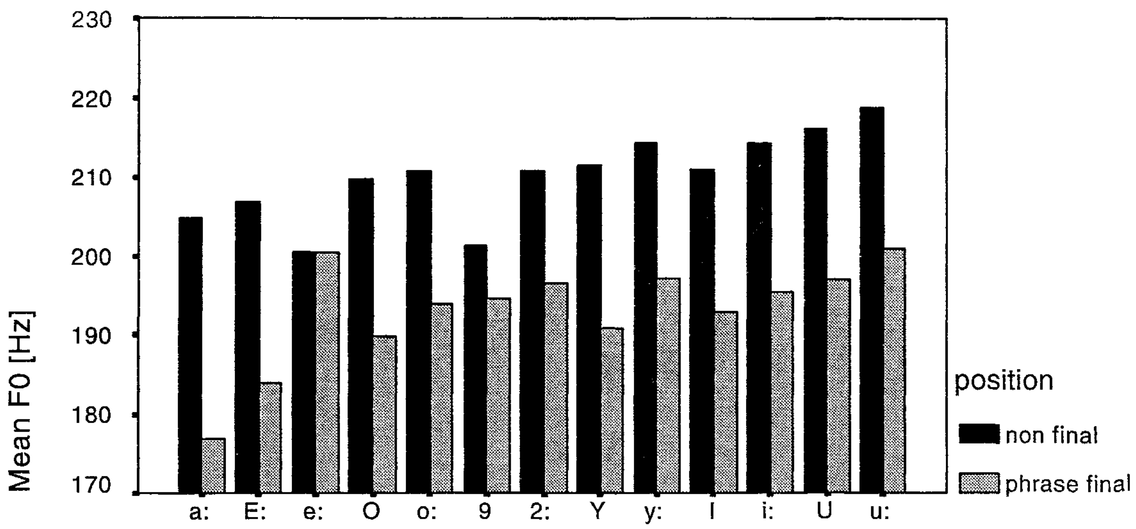


**Fig. 1:** Mean F0 values for vowels appearing in different macroprosodic contexts. The SAMPA-transcription is used.

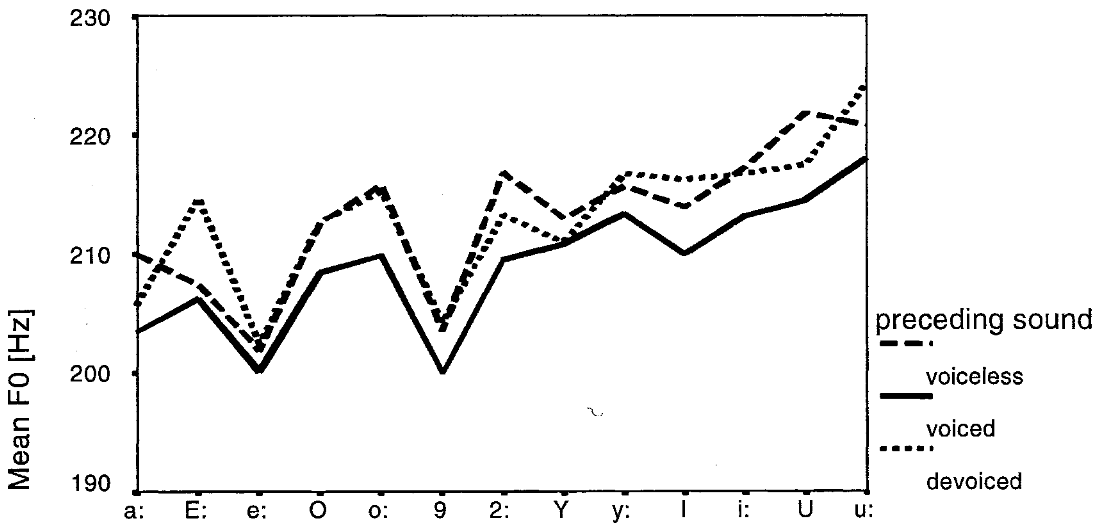


**Fig. 2:** The influence of voicing of the preceding consonant on vowel-$F_0$

*3.1.2.3 nasality*

Nasality of a preceding consonant affects vowel $F_0$ by a decrease of 2 Hertz on the average. Again, this tendency is less important for closed vowels than for open vowels. For French, Di Cristo [4] had stated a similar effect for two speakers, and found it reversed by a third one. He concluded that there is no influence at all. We found the influence to be significant, but the quantitave variation is probably too small to be of perceptual relevance.

*3.1.3 Intrinsic $F_0$ - values of consonants*

Unlike other studies of intrinsic $F_0$ in German ([1], [2], [3]), we decided to regard not only vowels but also voiced consonants. Fig. 3 gives a survey of the results.
As often stated ([6], [4]) voice bars during the occlusion of voiced plosives show a clearly lower $F_0$ than the other sounds. Further, there is an influence of the place of articulation on intrinsic $F_0$: the more advanced the place of obstruction is, the higher is $F_0$. These findings contradict those of Di Cristo [4] for French who has stated that retracting the tongue raises $F_0$.

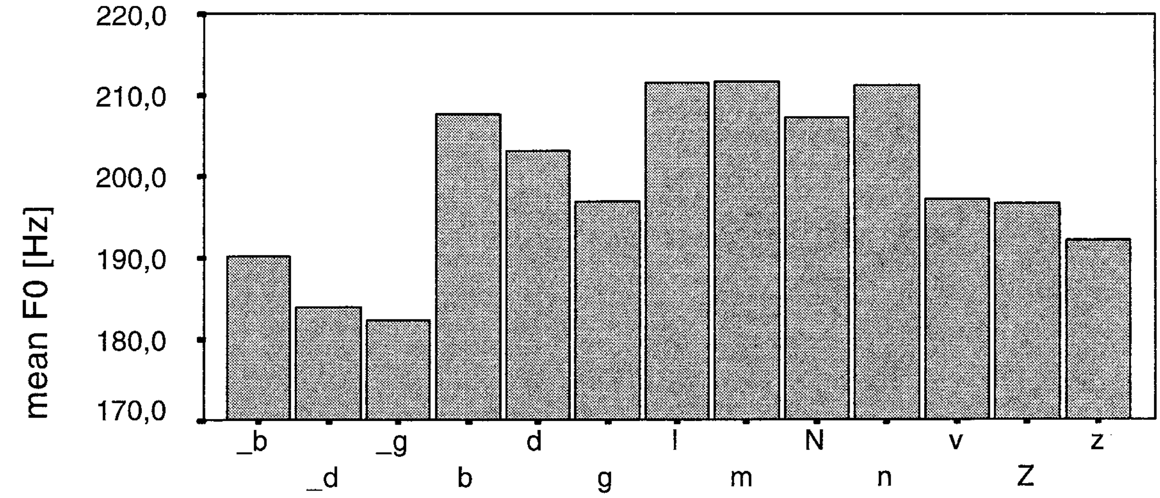


**Fig. 3:** Mean $F_0$-values for voiced consonants. Underscores mark voice bars in voiced plosives

**3.2 Duration**

3.2.1 Intrinsic duration values of vowels

The mean values for each vowel were calculated depending on their position in the carrier phrases. Apart of the fact that phonologically long vowels really have longer durations than short vowels, it can be shown that open and closed vowels behave differently. The open vowels [a:] and [E:] show the biggest differences between the durations of phrase-final and non-final realizations. However, the finding described by several authors (e.g.[2], [4], [7]) that open vowels are longer than closed ones in our data only applies for the vowels in final position. The data for the non-final positions even show the shortest realizations for [a:] and [E:] (see fig. 4). This may be a speaker-dependent variation or a corpus effect and has to be reexamined with more material.

3.2.2 Influence of the segmental context

*3.2.2.1 Voicing*

Voicing of a consonant causes an elongation of the following vowel. The values for devoiced consonants lie between those for voiced and those for voiceless consonants. These differences are significant, but the numerical effect is small (about 8 ms in average). This influence of voicing of the consonantal envinronment on vowel duration has been stated by several authors, see [4] for a survey.

*3.2.2.2 Manner of articulation*

The longest durations are found for vowels preceded by a sonorant. Vowels preceded by a plosive are about 3 ms shorter on the average. The vowels preceded by a fricative another 3ms shorter. The shortest vowels are those preceded by another vowel. Although these differences are significant, it is questionable if they are of any perceptual importance. The following consonants show less influence; the longest vowels are those followed by plosives.

*3.2.2.3 Place of articulation*

There is a tendency that an advanced place of articulation of a consonant increases the duration of a following vowel. The vowels preceded by a labial are the longest ones, smaller durations are found for vowels preceded by labiodental, alveolar and palatal sounds; vowels preceded by a velar are the shortest (see fig. 5). Antoniadis & Strube [9] found the longest values for vowels embedded in an alveolar context, whereas House & Fairbanks [7] claim that there is no influence at all. However, only the difference between the vowels preceded by labials and velars which is about 20 ms on the average is significant. The results of Peterson & Lehiste [8] show a contradictory effect for consonants following a vowel; i.e., vowel duration is proportional to tongue retraction.

3.2.3 Intrinsic duration values of consonants

3.2.3.*1 Voicing*

The occlusion phases for voiceless plosives are generally longer than for voiced ones, this is a well known fact. The devoiced pauses are shorter than the voiceless ones, this might be a cue for the perception of devoiced plosives.
Voiceless fricatives show the longest durations as well, but the shortest durations are found for the devoiced sounds. An explication for this irregular behaviour of devoiced obstruents might be that devoiced sounds are phonologically voiced, so they must be acoustically different from voiceless sounds only.

*3.2.3.3 Influence of adjacent vowels*

Neweklowsky [2] had found an influence of the following vowel on the duration of the explosion and the aspiration of [p]. We have studied this phenomenon for other plosives as well as for fricatives. Neweklowsky found the longest aspirations preceding [U], the shortest aspirations preceding [a] and values in between for [I].
Our data support the statement for plosives and fricatives preceding short vowels. However, for long vowels, the

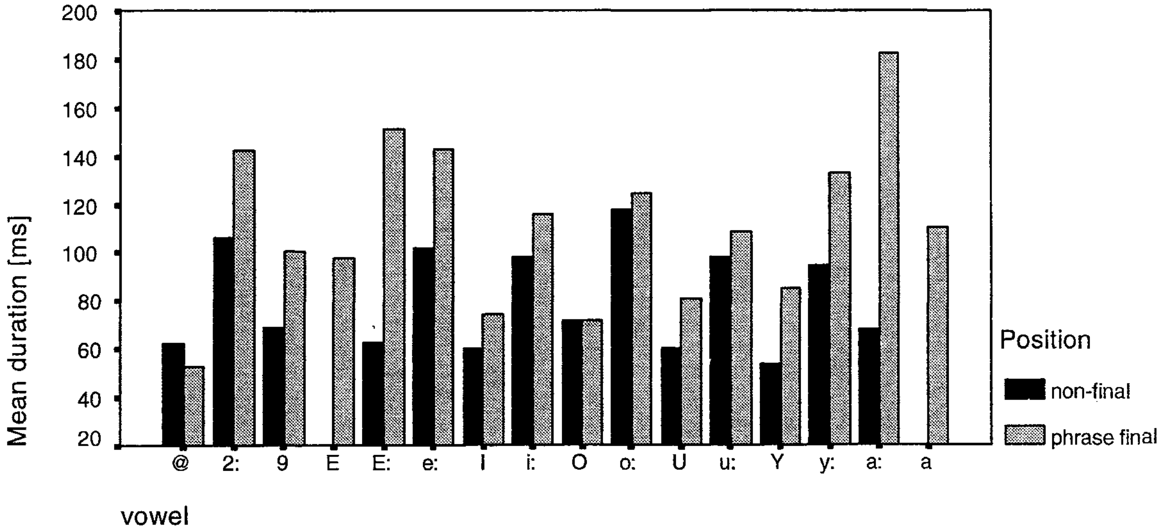


**Fig. 4:** Mean durations of vowels in phrase-final and non-final positons. (SAMPA-transcription)

longest durations of fricatives and aspirations are found preceding [i:] and [y:], whereas [u:] does not behave differently from [a:] (see fig. 6).

## 4. DISCUSSION

Our results about intrinsic prosodic values support other reports on this subject. We additionally found different influences of the segmental context on $F_0$ and duration. Most variations caused by context are quantitatively small, so it will probably not be necessary to take them into account in the generation of synthetic prosody. However, this is no more than a preliminary study, because it was carried out for one speaker only. We will obtain more reliable results in future investigating other synthesis inventories.

## REFERENCES

[1] Antoniadis, A. & Strube, H.W. (1981) : Untersuchungen zum "intrinsic pitch" deutscher Vokale. *Phonetica* **38**: 277-290

[2] Neweklowsky, G. (1975): Untersuchungen zur spezifischen Dauer und Tonhöhe der Vokale. *Phonetica* **32** : 38-60

[3] Möbius, B.; Zimmermann, A.; Hess, W.(1987): Microprosodic fundamental frequency variations in German. *Proc. of the XIth ICPhS*, p. 146-149 Tallinn, Estonia

[4] Di Cristo, A. (1985): *De la microprosodie à l'intonosyntaxe.*Aix-en-Provence

[5] Portele, T. (1994): Ein phonetisch-akustisch motiviertes Einheiteninventar für die deutsche Sprachsynthese. (Dissertation, Univ. Bonn)

[6]Lea, W. (1980): Prosodic aids to speech recognition. In: W. Lea (ed.) Trends in Speech Recognition

[7] House, A.S. & Fairbanks, G. (1953) The influence of consonant environment upon the secondary acoustical caracteristics of vowels. *JASA* **25**: 105-113

[8] Peterson, G.E. & Lehiste, I. (1960): Duration of syllable nuclei in English. JASA 32: 693-703

[9] Antoniadis, Z. & Strube, HW. (1984): Untersuchungen zur spezifischen Dauer der deutscher Vokale

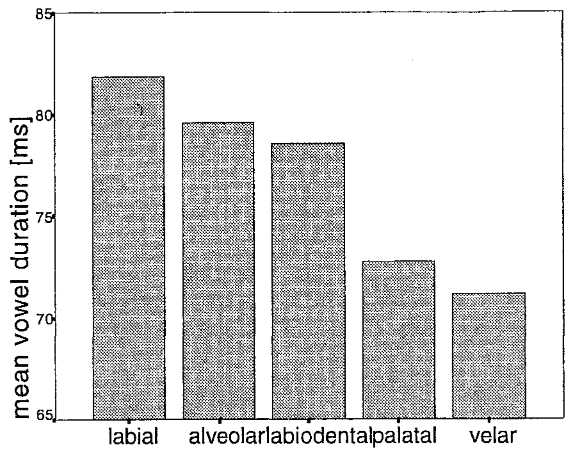


**Fig. 5:** Duration of vowels depending on the place of the preceding consonant

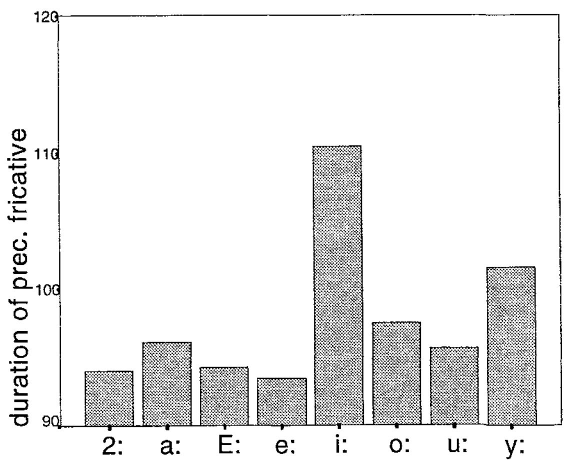


**Fig. 6:** Duration of fricatives depending on the following vowel