DURATION, F-PATTERNS AND TEMPO IN GERMAN SYLLABLES

Duncan Markham, Dept. of Linguistics, Faculties,
Australian National University.

ABSTRACT: Segmental duration and vowel formant frequencies in German nonsense syllables at two different speeds were investigated The results are compared to previous research in this area and some modelling issues are discussed.

INTRODUCTION

Investigations into the effects of tempo on syllable and segment duration have been more concerned with measurement and the behavioural characteristics of the syllable coda, largely ignoring the onset or pre-vocalic element (Port 1977). Some major work in this area has also neglected intraspeaker variation (Crystal and House 1988a,b), whilst durational modelling seems to be based on a production-independent premise (Klatt 1976; Port 1981). Little regard is given to the durational proportion accounted for by a segment in a syllable (O'Shaughnessy 1981; and above works). In addition, little work has been done on German, especially in regard to modelling.

This paper looks at the duration, degree of reduction, and proportional characteristics relative to total utterance duration of all segments in the syllable. Formant frequencies and the effects of tempo on them are also described, relative to previous investigations of vowel undershoot in German (Delattre 1981) and Swedish (Lindblom 1963; Engstrand 1988).

Lindblom (1963) investigated the effects of tempo and stress on formant frequencies and vowel duration in Swedish and attempted to derive a model from the results. He found that vowel undershoot and duration are directly correlated, and proposed that undershoot was an effect of speech organ failure — the physiological inability of articulators to reach a target position at given speeds. Delattre (1981) investigated the effects of stress on vowels in German. By contrasting stressed and unstressed syllables in pairs of real words of similar form he obtained data for F1 and F2 for each vowel in the given languages in order to give an indication of vowel reduction. He found by plotting F1 against F2 (simulating the vowel trapezoid) that unstressed vowels were never located more peripherally than their stressed counterparts. However, his data did not show a universal cross-formant movement towards a central, or schwa, value. Whilst one formant might move "towards" schwa, there were instances where the other formant becomes more peripheral (though never more peripheral than the extreme for stressed vowels). Engstrand (1988) investigated the articulatory correlates of stress and speaking rate for three Swedish vowels /i a u/, and measured segmental durations and vowel formant frequencies across two rates and two levels of stress. In the utterance *seja pVCVp ijen*, the informants were instructed to stress *pVCVp* for both rates, and then *ijen* for both rates. Engstrand found no significant change in formant frequencies for rate.

THE EXPERIMENT

An experiment was constructed to test the effects of tempo on the duration and F-pattern of nonsense syllables with the structure /bVt/, where V could be any of the 15 vowel phonemes in Standard German. A male native speaker of a dialect of south-western Germany (Schwäbisch) repeated each syllable three times at a tempo chosen by him, and then in a second reading was instructed to read the syllables at a faster tempo. (The resultant syllable rates were : NORMAL ≈ 1syl/sec; FAST ≈ 2.4syl/sec.) Wide-band spectrograms were made and a total of 66 tokens representing 12 phonemes was measured from the release of the initial stop /b/ to the release of the final closure for /t/. Mid- or peak (where the sign of the formant gradient changed from positive to negative, or vice versa) frequencies of the first three formants were measured. Measurement error was maximally ±2.5ms/±20Hz.

Results

LONG and SHORT refer to phonemic vowel length. Ø and + refer to NORMAL and FAST tempo.
It was found that the mean duration of VOT for the initial stop did not vary greatly across tempo or vowel environment, and in three cases formed an almost constant proportion of the total mean utterance duration. A straight line of best fit was calculated for each of the four pairs of duration and proportion

means for VOT. It was found that the minimum value for the correlation coefficient *r* was *r=0.981* and the maximum was *r=0.99*, indicating a near perfect fit. Where *y*, the dependent variable, represented the proportional duration, and *x*, the independent variable, was the absolute duration, it was found that some regularity obtained for the intercepts and gradients of the four lines of regression.

	LONGø	LONG +	SHORTø	SHORT+		LONGø	LONG +	SHORTø	SHORT+
/b/	n=14	n=26	n=15	n=13	**/b/**	n=14	n=25	n=15	n=12
Xms	35.1	35.6	34.1	30.1	*X%*	8.50	12.34	11.77	12.52
sd	12.6	16.2	12.7	16.6	*sd*	2.83	5.15	4.13	5.52
/V/	n=14	n=26	n=15	n=13	**/V/**	n=14	n=25	n=15	n=12
Xms	245.7	165.6	112.0	101.5	*X%*	59.79	56.98	38.81	41.66
sd	17.7	15.9	12.7	15.2	*sd*	2.27	5.00	3.48	4.82
/t/	n=14	n=25	n=15	n=12	**/t/**	n=14	n=25	n=15	n=12
Xms	130.3	89.2	142.2	110.3	*X%*	31.71	30.69	49.42	45.82
sd	13.1	15.1	13.2	11.2	*sd*	2.44	4.19	4.88	3.20
TOTAL	n=14	n=25	n=15	n=12					
Xms	411.1	290.4	288.3	241.7	%	100.00	100.00	100.00	100.00
sd	28.6	24.5	15.7	29.5					

Table 1. Durational and percentage means and standard deviations for all segments and TOTAL across tempo and phonemic vowel length.

	LONGø	LONG +	SHORTø	SHORT+
INTERCEPT(b)	.772	1.048	.826	2.694
GRADIENT(a)	.22	.312	.321	.317
r	.984	.982	.99	.981

Table 2. Intercept, gradient and correlation coefficient values for lines of best fit predicting percentage of total utterance occupied by /b/.

NORMAL values show similar intercepts, whilst those for FAST are somewhat higher. In addition the slopes of all except LONGø are very close indeed. An analysis of correlation between proportional duration and absolute duration, with the latter as the dependent variable, *y*, naturally yields the same *r* values, but slope and intercept of the lines of regression will be different.

	LONGø	LONG +	SHORTø	SHORT+
INTERCEPT(b)	-2.257	-1.915	-1.834	-7.009
GRADIENT(a)	4.4	3.089	3.055	3.036

Table 3. Intercept, gradient and correlation coefficient values for lines of best fit predicting duration of /b/ from percentage of total utterance.

Using knowledge of the above equations an attempt was made to predict values for the total utterance duration (TOTALpred) from the VOT duration, and vice versa. Thus, using the equations generated from Table 2 (predicting proportion from VOT), TOTALpred was obtained, and using Table 3 (predicting VOT from proportion), VOTpred was obtained:

Eq.1: $\text{TOTALpred} = \frac{100\text{VOT}}{b + a.\text{VOT}}$ Eq.2: $\text{VOTpred} = \frac{100a.\text{VOTpred}}{\text{TOTAL}} + b$

VOTpred was found to be unacceptable as an approximation of VOT. For all environments VOTpred deviated radically from the observed values. TOTALpred provides a somewhat better (though by no means acceptable) estimation of TOTAL. The mean difference between TOTAL and TOTALpred ranged from 11.13ms for SHORTø to 20.35ms for LONGø. From these results one might conclude that VOT plays a part in determining the total utterance duration, and not that total utterance duration determines VOT. If VOT duration corresponds to a specific durational template for the whole utterance, then it falls on the last two segments to realise this pattern. Lack of articulatory precision may result in some durational overshoot or undershoot, thereby providing a possible reason for failure in predictive accuracy. Alternatively, further manipulation of the proposed formulae might well yield better predicted values. It is clear from these results that, although the lines of regression showed very good fits, the small deviations from the lines shown by some points had considerable effects on the predicted values, where minute deviations in the input lead to considerable and unacceptable errors in prediction.

Unlike the data for /b/, no significant durational correlations could be found for the coda /Vt/. Of course, there is an approximately inverse relationship between the two segments due to the apparently non-random durational values for VOT and the total duration. Thus, where [V] deviates from the mean for /V/, [t] must show an inverse deviation. Analysis of correlation for /V/ and /t/, /V/ and TOTAL, and /t/ and TOTAL yielded generally low (*r*<0.65) *r* values.

The above results differ considerably from those of Kohler et al (1981), who found that all segments in a number of nonsense words formed a regular proportion of the total utterance. It should be noted however, that Kohler's data were elicited at an externally controlled tempo, whereas in my research, tempo was determined by the informant, thereby permitting greater temporal variation.

As can be seen in Table 1, LONGø vowel duration is 245.7ms, whilst LONG+ vowel duration is some 80ms shorter at 165.6ms. The difference between SHORTø and SHORT+ however is only 10.5ms. Clearly, this very small degree of reduction for short vowels seems to be indicative of vowel incompressibility — the inability of vowels to shorten beyond a certain point (often referred to as the durational minimum (cf Klatt 1975)). These contrasts can be expressed more effectively via a reduction ratio α, as shown in Table 4.

	LONGø~LONG+	SHORTø~SHORT+	LONGø~SHORTø	LONGø~SHORT+	LONG+~SHORT+
/V/	0.67	0.91	0.45	0.41	0.61
/t/	0.68	0.77	1.09	0.85	1.24
TOT	0.71	0.84	0.70	0.59	0.83

Table 4. α (reduction ratio) values for /V/, /t/ and TOTAL.

The figures for vowels in the first, third and fourth columns show that reduction across tempo and phonemic length is not constant. As the utterance gets shorter, the vowel compresses less and less. Thus, LONG+, SHORTø, and SHORT+ are to LONGø 0.67, 0.45 and 0.41 respectively, whilst for TOTAL the values are 0.71, 0.70 and 0.59. The vowel clearly reduces from LONG to SHORT, but reduction beyond this is very small indeed, despite the increase in tempo.

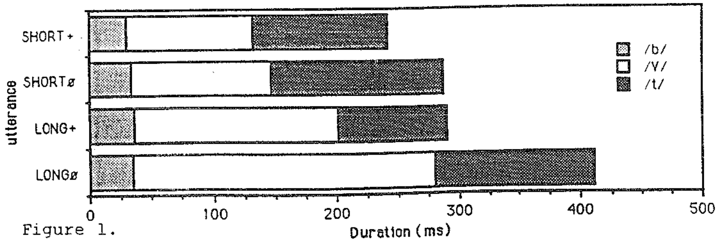

Figure 1.

In marked contrast to this is the fact that the total duration for LONG+ and SHORTø is virtually identical. This can be explained immediately by an examination of the /t/ durational means in Table 1. For short vowels the duration of /t/ is considerably larger than for long vowels. Proportional figures given in the second half of Table 1 demonstrate this, and Figures 2 and 3 provide a graphical illustration of the data.

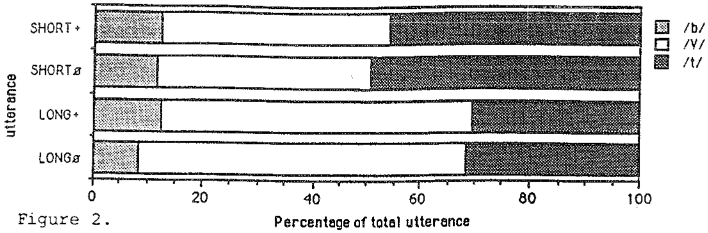

Figure 2.

For LONG, /t/ forms approximately 31% of the total utterance, whilst for SHORT the value is much higher — 45-49%. As /t/ still reduces by a fairly large amount from SHORTø to SHORT+ (α=0.77) but the vowel cannot, the mean percentage of the utterance occupied by the vowel actually increases from 38.81% to 41.66%. Notably, however, the tempo reduction for SHORT is not as large as for LONG (α=0.68). I see only two possible explanations for this — 1: that consonants also have a minimal duration (though I know of no research confirming this), or 2: that the production mechanism seeks to maintain constant proportional values for all segments across rate. Under the second hypothesis /t/ reduces less in order to compensate for the increase in the proportion potentially occupied by /V/, but cannot attain the same proportional value as for SHORTø, because this would result in the difference between TOTALø and TOTAL+ being in the order of 27ms, rather than the observed 46ms (Fig. 4), presumably generating a perceptually aberrant token for the given speech rate. Hypothesis 2 directly contradicts, and provides an intuitively more attractive explanation than that suggested above, as a result of the modelling attempt.

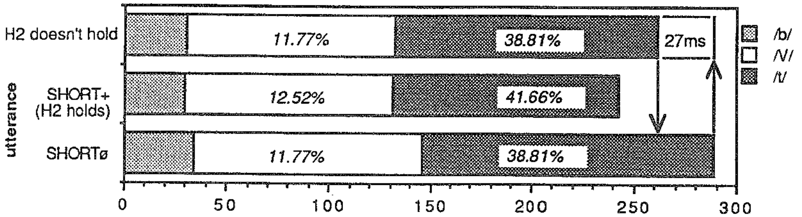

Fig. 4. Durational and percentage values generated if Hypothesis 2 does and does not hold.

There are obviously target percentages for the production mechanism, considering the fairly low values of standard deviation (sd<6ms) and, more to the point, target durational values for segments (11<sd<18) and TOTAL. Standard deviations for all segments, irrelevant of duration, fall within a certain range, perhaps indicating that for any segment there is a specific deviational tolerance. The only obvious explanation for such a constant range would have to relate to articulatory constraints common to all segments. No relationship was observed between vowel height and duration, which, considering the small size of the corpus, is not particularly surprising.

Whilst I have proposed a predictive model for the syllable onset (VOT), I cannot do the same for the syllable coda. The proportional data provide the beginnings of such a model for the whole syllable, but the lack of predictive tolerance means that more data would need to be collected in order to proceed further. Previous predictive models (Port 1980; Klatt 1975; Lindblom and Rapp 1973) have sought to generate segment durations solely using durational data, with no reference to proportional data, and no indication of the tempo to which this modelling is appropriate. These models are concerned solely with the absolute duration of a given segment (usually vowels) in their various syllabic and sentential contexts. They ignore the contribution of all segments to the syllable and word, and how this changes with an increase in segments, syllables and words, and with changes in intonation. O'Shaughnessy (1981), for instance, provides a wealth of data from which segment durations in French may be derived via a multiplicative approach, but ignores the syllable.

F-Patterns

Figure 4 shows F-pattern means and reduction for all vowels, whilst Figure 5 compares these values with those of Delattre (1981) for F1 and F2 of stressed and unstressed vowels in German. Non-low unrounded front vowels showed similarity in *degree* of undershoot (change in value of either formant), though their absolute formant-frequencies differed considerably. The formants of the rounded front vowels /ʏ ø/ showed more similar frequencies and the latter vowel reduced to a similar degree to Delattre's figures. Non-low back vowels /u o/ were almost identical to Delattre's figures at NORMAL tempo, whilst /ʊ/ had a slightly lower F1. Reduced values were different in the nature of undershoot (Delattre's figures show change primarily in F2, whereas mine change in both formants or F1) but again the degree was similar. Markedly different was the reduction of /a ɑ/. Whilst NORMAL frequencies were almost identical, Delattre's figures for reduction showed a considerable decrease in F1, whilst my figures show significant change solely in F2.

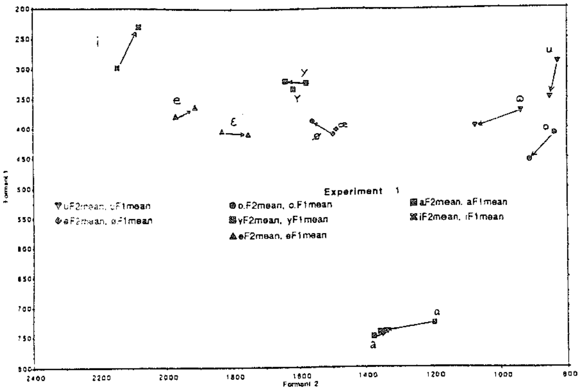

Figure 4.

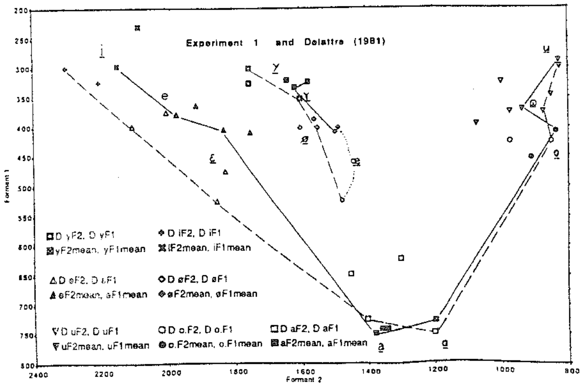

Figure 5.

It is exceedingly interesting that, despite the fact that Delattre's informants were almost certainly speakers of Standard German, my informant, who spoke with considerable dialectal colouring of his speech, showed similar un-reduced formant frequencies in at least six of the vowels measured. Many differences obtained for the reduced values, though general trends could be seen for both sets of data. I cannot comment on the significance of these differences, as the reduced values in the two studies are functions of different processes (stress versus tempo), which may well have different effects (though few researchers seem to have qualms about not differentiating between types of reduction).

The formant data were also plotted against vowel duration, and for most formants of most vowels there seemed to be a strong correlation between change in duration and change in formant frequency, as found by Lindblom (1963). At the very least this shows that stress (which was not examined) does not play the definitive rôle in determining reduction, as argued by Gay (1978) and Engstrand (1988).

REFERENCES

Crystal, T.H. and House, A.S. (1988a) "Segmental durations in connected-speech signals: current results", J.Acoust.Soc.Am. 83, 1553-1573.

Crystal, T.H. and House, A.S. (1988b), "Segmental durations in connected-speech signals: syllabic stress" J.Acoust.Soc.Am. 83, 1574-1585.

Delattre, P. (1981), *Studies in comparative phonetics*, (Julius Groos: Heidelberg)

Engstrand, O. (1988), "Articulatory correlates of stress and speaking rate in Swedish VCV utterances.", J.Acoust.Soc.Am. 83, 1863-1875.

Gay, T. (1978), "The effect of speaking rate on vowel formant movements", J.Acoust.Soc.Am. 63, 223-230.

Klatt, D.H. (1976), "Linguistic uses of segmental duration in English: acoustic and perceptual evidence", J.Acoust.Soc.Am. 59, 1208-1221.

Kohler, K.J. et al (1981), "Sprechgeschwindigkeit in Produktion und Perzeption", AIPUK 16, 139-205.

Lindblom, B. (1963), "A spectrographic study of vowel duration", J.Acoust.Soc.Am. 35, 1773-1781.

O'Shaugnessy, D. (1981), "A study of French vowel and consonant durations", J.Phonetics 9, 385-406.

Port, R.F. (1977), *The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words.* (IULC: Bloomington)

Port, R.F. (1981), "Linguistic timing factors in combination", J.Acoust.Soc.Am. 69, 262-274.