

TIME AS A FACTOR IN THE ACOUSTIC VARIATION OF SCHWA

William J. Barry

University of the Saarland, Germany

ABSTRACT

The question of schwa as a "targetless" vowel is discussed in connection with the methods employed in earlier studies to address the issue. It is argued that the evidence for the assumption of targetlessness is so far unconvincing. An experiment is presented in which schwa is produced in symmetrical vowel and consonant contexts under varying speech rate conditions. It is shown that the contextual influence on schwa depends on its duration, a result that is incompatible with the concept of a targetless vowel. This conclusion is discussed in relation to the possible phonological status of the target specification

1. INTRODUCTION

In the past several years a number of acoustic phonetic studies in three different Germanic languages [1-4, 6] have implicitly or explicitly taken up Browman & Goldstein's 1992 articulatory discussion [5] of the vocalic element schwa, as a possibly "targetless" vowel, i.e., as completely dependent on its surrounding consonantal and trans-consonantal vowel context. Using different parameters and different statistical approaches, Bates [3] and van Bergem [4] were both able to show a high level of explained acoustic variance in schwa, and interpreted these results as evidence supporting the targetless hypothesis. Barry [1,2] showed the proximity of the schwa to the mid-point of some speakers' F1-F2 vowel space, a finding that can be interpreted either as the product of the combined contexts or, alternatively, as a neutral target position. Comparison of German speakers from different regional backgrounds showed a case of lexical schwa with a clearly context-independent target.

The hypothesis of a targetless lexical-schwa thus cannot be maintained without exception. In both [3] and [4], the conclusion of targetless is based on the high level of explained variability in schwa - primarily in F2 - (92% and 73% for F2 and F1, respectively in [3] and 71-77% and 29-38% for F2 and F1 [4]. These differences in the level of explained variability themselves suggest that procedural factors or factors connected with the speech material examined may have a considerable effect on the results. Indeed, Koopmans van Beinum (1994) shows that schwa decreases systematically in its degree of variability from spontaneous speech, through a read version of the spontaneous text, to a carefully produced nonsense corpus recorded for diphone synthesis. Close consideration of the parameter definitions used in Bates [3] and v. Bergem [4] allows for further scepticism towards the evidence supporting the targetlessness hypothesis.

Firstly, in Bates [3] right- and left-hand consonantal context (and, incidentally, duration too) were considered as *independent* variables, with F1 and F2 at the schwa mid-point as *dependent* variables. The quantification of the context was based on F1 and F2 measurements at the onset and offset of schwa. Given the short duration of schwa - 60 ms is a liberal estimate - the high level of combined dependency of the schwa values (measured at the mid-point) on the onset and offset values is not surprising.

In v. Bergem [4], by contrast, right- and left-hand consonantal context *and* the vowel preceding or following the schwa-syllable were considered as independent variables. Duration together with schwa-onset, -offset and -midpoint were the *dependent* variables. The consonantal context was differentiated merely by phonemic category, not by any quantification as it was in Bates' study.

In summary, there are sufficient methodological differences between the two studies - ranging from the measured parameters, through the definition of dependent and independent variables, to the statistical evaluation procedures applied - to expect differences in the degree of explained variability. But there are also aspects about the *interpretation* of the results that are problematical. Both studies come to the conclusion that schwa is targetless due to the degree of variability explained by the contextual factors. However, percent explained variability only captures the *degree of systematicity* in the variation; it says nothing about the *extent* to which the dependent variable is affected. Figures 1 illustrate this:

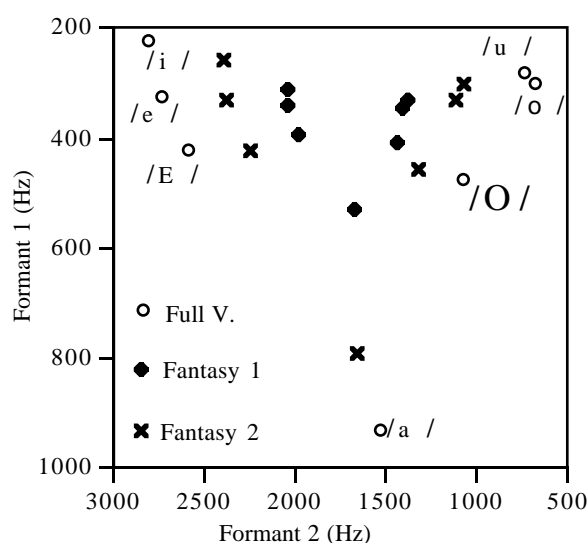


Fig.1 Fantasy Schwa with Constant Regression Values

The two formant charts in the figure showing a smaller and a larger spread of fictitious schwas as a function of context vowels give same degree of explained variability (71% for F2) despite the totally different strength of contextual "pull" exercised by the context. In conclusion then, systematic variation of either F1 or F2 is, in itself, not a valid argument in support of or against the purported targetlessness of schwa.

Finally, some aspects of the articulatory data provided by Browman & Goldstein [5], the study referred to in both [3] and [4], indicate that schwa does have an articulatory target, albeit a weakly defined one. In that study, targetlessness is investigated with reference to tongue dorsum gestures, where the dependency on the pre- and postschwa *vowel*, not the consonantal context, is at issue. There is the assumption, common to many views of speech production, that the tongue body is controlled as a vocalic subsystem, moving from vowel to vowel with consonantal gestures superimposed. Thus, Bates' [3] study is not related to this level of articulatory control; she investigates only the acoustic reflex of the *consonantal* gesture at the beginning and end of the acoustic vocoid segment. Van Bergem [4] includes one vowel, pre- or postschwa in his independent variables, but does not provide a bilateral vocalic context, and consequently also pays more attention to the consonantal context. Neither study takes up the issue which, ultimately, for Browman & Goldstein, provided evidence against the targetlessness of schwa, namely the fact that in symmetrical vowel contexts, particularly the extremes (e.g. /i: __ i:/, /a: __ a:/ or /u: __ u:/) there was a clear movement of the tongue body away from the position demanded by both flanking vowels.

For the experiment reported here it was hypothesized that if a vowel target does exist, it will be approximated to a greater or lesser degree depending on the amount of time available. A symmetrical pre- and post-context was used to provide data which is directly interpretable in terms of context independence. Phrases containing schwa between symmetrical consonantal and vocalic contexts were used, read at slow, normal and fast speeds. Schwa F1 and F2 measurements are examined in relation to the flanking stressed vowel categories, and in relation to the place of articulation of the flanking consonants.

3. ANOTHER LOOK AT TARGETLESSNESS

Despite the undeniably strong influence of vocalic and consonantal context on the realisation of schwa, there should be some acoustic reflex of the tongue-body movements away from flanking vowels noted by Browman & Goldstein in symmetrical contexts. The question should, therefore, not be directed towards the negative expectation of targetlessness, and linked to the fact of a high degree of explained schwa variability, but should be formulated positively: *Are there any indications of schwa independence?*

Symmetrical contexts, which were included in Browman & Goldstein's data, and different rates of speech, as implied in Koopmans v. Beinum's different styles of production data,

offer conditions under which movements towards a possible target position may be revealed. To this end, phrases were devised which provided symmetrical vowel and consonant contexts for schwa. The consonants were labial, alveolar and velar stops (/b, d, g/), and the vowels were high, high-mid and low-mid front unrounded and back rounded, and central open German vowels (/i, e, E, a, O, o, u /). All symmetrical combinations of vowels and consonants were used ($7 \times 3 = 21$ stimuli):

E.g. "die liebe Bine", "rigide Dielen" "ich biege Giebel"
... "ich lobe Bozen" "die Mode Dosen" "im Soge gotisch", etc. Explanatory, though sometimes far-fetched sentences were devised to provide a possible wider context for the phrases.

The 21 phrases were read 5 times in different random orders at each of three speeds: normal, slow, fast (in that order), by four speakers (2F, 2M) from different regions of Germany, but all speakers of High German.

Tape recordings were digitised at 10 kHz sampling frequency on a Kay CSL workstation. Phrase, pre- and post-schwa context-vowel and schwa durations were measured using synchronised cursors on an appropriately zoomed time-pressure waveform and the corresponding spectrogram. Formants one and two were measured on a 14-order LPC spectrum calculated over a 30 ms window located round the vowel mid-point. Statistical analyses were performed using the 7.5 Windows95 version of SPSS.

4. RESULTS

Each of the four speakers produced the phrases at systematically different articulation rates for the three tempo conditions. However, the degree to which they varied the rate was very different (Fig. 2).

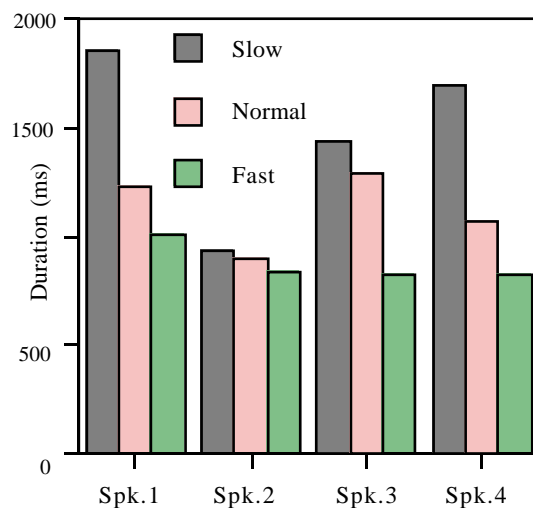


Fig.2 Phrase duration as a function of speech rate

Nor was the variation in schwa duration in a clear relation to the phrase-level articulation rate (Fig. 3) Speaker 3 produced

the most evenly spaced schwa modification across tempo, while speakers 1 and 4 differentiated very clearly between slow and the other two conditions. Speaker 2 was basically unable to speak slowly at all. After group treatment, the inter-speaker differences will be considered again.

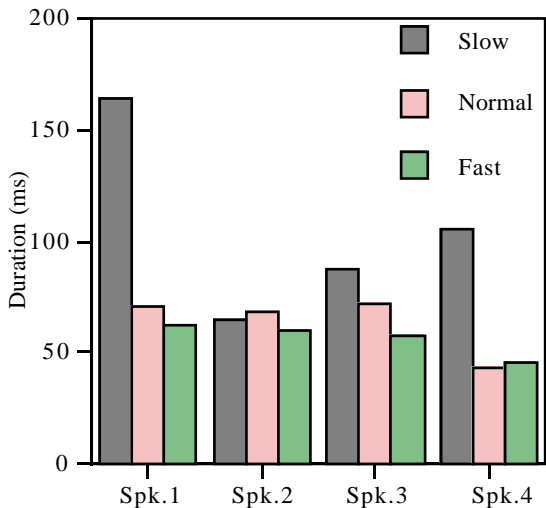


Fig.3 Schwa Duration as a Function of Tempo

The well-documented effect of consonantal context on the schwa values, is confirmed in the present data. In addition, a clear dependence on speech rate is observable (Fig. 4). Four consonantal contexts are defined in the figure for the three places of articulation because the velar context is divided into front-vowel and back-vowel variants. This reflects the shared tongue dorsum as primary articulator for both vowel and velar consonant production. The importance

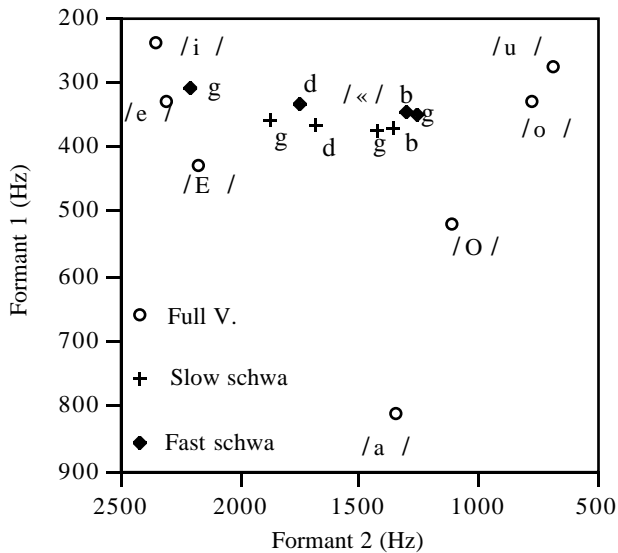


Fig.4 Slow and Fast schwa in Consonant Context

of the shared articulator is also apparent in the tempo effect. The slow-to-fast shift of F2 is much stronger for both velar

variants than for the labial and alveolar contexts. There is a systematic though much less dramatic effect of tempo on F1, the lower values for the fast rate presumably reflecting a generally reduced degree of opening.

The vowel context effect, focussed on in Browman & Goldstein's [5] articulatory study, but also confirmed in v. Bergem's [4] acoustic analysis is clearly present in the present data. And again, there is a clear increase in the extent of the vocalic context with increased speech rate (Fig. 5).

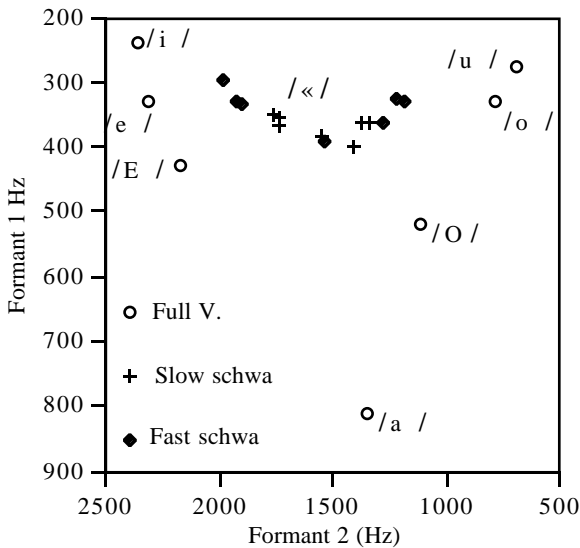


Fig. 5 Slow and Fast Schwa in Vowel Context

In this case, the increased contextual effect applies to both Schwa-F2 and -F1, whereby the reduced mouth-opening appears not to affect the open vowel /a:/ context.

The complementarity of the consonantal and vocalic effects is observable in a comparison of correlations between Schwa-F2 and the individual Context Factors with *and* without control for the second factor. The reduction in explained variability for each of the factors controlling for the other is a mere 1%, indicating a very low level of covariance.

	r	Expl. var.	r (contr.)	Expl. var.
Cons. Ctxt:	0.637	40.5%	0.628	39.4%
Vowel Ctxt:	0.622	38.6%	0.612	37.5%

A multiple regression on the F2 data showed that 62.8% of the overall F2 variance was attributable to the consonantal and vowel context categories (corrected $R^2 = 0.628$). Carried out separately on the slow, normal and fast conditions, the degree of explained variability increased only slightly from slow to fast, as is to be expected with the degree of systematicity basically independent of the strength of the influence.: Slow: $R^2 = 0.620$ < Normal: $R^2 = 0.648$ < Fast: $R^2 = 0.681$. Interestingly, the standardised Beta coefficients, reflecting the relative weighting of the consonantal and the vocalic context factors shifted from a clear dominance of the consonantal context in the slow condition to a dominance of the

vocalic factor in the fast condition:

Slow: Cons: 0.533 > Vowel: -0.457

Normal: Cons: 0.526 > Vowel: -0.489

Fast: Cons: 0.507 < Vowel: -0.533

Within a conceptual framework of separate control sub-systems for vowels and consonants, an increase in the relative weight of the vowel-context factor is to be expected given the reduced temporal separation of the vocalic events as speech rate increases.

To examine the extent of F1 and F2 change as a function of speech rate four-way ANOVAS were performed (tempo x consonant context x vowel context, with speaker as random variable). In the case of F2 the expected strong main effects for consonant context (F df3, = 31.8 $p < 0,001$) and vowel context (F df6 = 43.8, $p < 0,001$) are found but, as expected, no main effect for tempo. However, there is a strong interaction, tempo x vowel context (F df12 = 4.2, $p < 0,001$), reflecting the increasing strength of the different flanking vowels with increasing tempo. The tempo x consonant context interaction, which would also be expected to be significant, failed to manifest itself (F df6 = 2.06, $p = 0,110$). However, a strong three-way interaction, speaker x tempo x consonant context (F df18 = 6.8, $p < 0,001$), betrays the reason: Speaker two differs from the other three, as his phrase duration and schwa duration patterns presented above might lead us to expect. The tempo x consonant context interaction for the other three speakers, both individually and as a group (F df6 = 8.6, $p < 0,001$), is highly significant, while speaker 2 shows no effect (F df3 = 1.2, $p = 0,301$). In terms of *vowel* context, however, he conforms to the pattern found for the other three speakers, and shows a significant tempo x vowel context interaction.

The picture with F1 differs considerably from that found with F2. A significant main effect of tempo might be expected since there is a systematic reduction of overall average F1: Slow: 369 Hz > Normal 348 Hz > Fast: 338 Hz. However, this is not borne out although, in interaction with vowel context, the tempo effect on F1 is significant (F df12 = 2.3, $p = 0,025$). The consonant context is also non-significant, either as a main effect (F df3 = 1.6, $p = 0,256$) or in interaction with tempo (F df6 = 1.2, $p < 0,351$).

5. DISCUSSION AND CONCLUSION

The results indicate that the contextual influence on schwa varies with the rate of speech. This provides support for the assumption that schwa is a vowel with a neutral target towards which the articulators move, not a completely targetless vowel. The more time at their disposal, the closer the articulators get to the assumed target. However, the result still leaves the phonological and – assuming some sort of link between the optimal phono-logical characterisation and the psychological representation – psychological status of schwa target open. If the "neutral" (i.e. mid-central) target implied by the results of the present study is to be seen as phonological, the complete non-specification of tongue-

position features cannot be seen as acceptable, however useful it may be at the level of formal system description. If, on the other hand, formal non-specification of schwa is maintained, there is still a plausible physiological basis for the "neutral target" suggested by the data.

It can be argued that a vocalic "slot" in the segmental tier may be underlyingly underspecified in terms of tongue position features without having to take over the neighbouring vowel specification. The arguments for maintaining a phonologically unspecified slot are considerable. Firstly, there is no convincing way to formally characterize the labial and alveolar consonantal influence on the schwa. The labial effect cannot be equated with [+round] nor can the alveolar influence be plausibly interpreted as [+front], which is the observed effect on schwa, since it is the pull of the tongue tip and blade (= [coronal]) which can be assumed to cause it. Secondly, the non-discrete nature of the influence militates against the simple spreading or linking of features, even those of the formally unproblematical vocalic context. This problem, of course, has been, and remains a fundamental problem of phonological explanations for gradient phenomena.

Thus, assuming that neither the consonantal nor the vocalic contexts impinge phonologically on the non-specified schwa skeletal slot, it can be claimed that a complex muscular sub-system such as the jaw-tongue-lip system serving vowel distinctions will move naturally towards a "relaxation" target during periods of non-specification. The target can be plausibly defined as the position determined by the muscular equilibrium established during acquisition of the system of stressable vowel targets. It is the muscular state within the "vocalic subsystem" from which the vowel targets can be reached most economically. This should be, *ceteris paribus*, the mid-point of the individual's articulatory vowel space, and consequently also close to the centroid of representative acoustic vowel measurements.

6. LITERATURE

- 1 Barry, W. J. (1995a): Schwa vs. schwa + /r/ in German. *Phonetica* 52, 228-235.
- 2 Barry, W. J. (1995b): Variation in schwa + /r/ in German. *Proceedings ICPhS95*, Stockholm, Vol. 2, 214-217.
- 3 Bates, S. (1995): Acoustic characteristics of schwa: A comparative study. *Proceedings ICPhS95*, Stockholm, Vol. 3, 230-233.
- 4 v. Bergem, D. R. (1994): A model of coarticulatory effects on the schwa. *Speech Communication* 14 (2), 143-162.
- 5 Browman, C. P. & Goldstein, L. (1992): "Targetless" schwa: an articulatory analysis. In: G. J. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II*, 26-56. Cambridge: CUP.
- 6 Koopmans van Beinum, F. J. (1994): What's in a schwa? Durational and spectral analysis of natural continuous speech and diphones in Dutch. *Phonetica* 51, 68-79.