Schwa out of Control?

Stefanie Jannedy, Susanne Fuchs and Melanie Weirich

Center for General Linguistics (ZAS Berlin)
E-mail: jannedy@ling.ohio.edu

Abstract

This study is a follow-up of our previous work and investigates transitional schwas between consonant clusters as a consequence of a gestural separation due to a slow speech rate. These uncontrolled schwas are compared to similar sequences with lexically specified schwas. Articulatory and acoustic data of 6 German speakers were recorded. Preliminary results provide evidence that transitional schwas fall out as a byproduct of a time delay between adjacent gestures but additionally need a low tongue back position.

1 General theoretical background

German, just as English, is characterized by the rhythmic alternation of strong and weak syllables. Weak or unstressed syllables contain short or reduced vowels like schwa which in some instances, can be the only difference between words that contain a lexically specified consonant cluster and those that do not. Examples in German are *geleiten* 'to accompany' that contrasts with *gleiten* 'to slide' or *beraten* 'to advise' that contrasts with *braten* 'to fry'. In some instances, under the influence of a faster rate of speech for example, weakening of the unstressed syllable nucleus is observed which can eventually result in the neutralization between such pairs of words. Weakening of the unstressed syllable nucleus in German has been described and explained in terms of a phonological deletion rule (Kloeke, 1982). In recent years however, alternative explanations based on gestural reorganization have been proposed for such observations (Kohler, 1990; Browman & Goldstein, 1990). A gestural reorganization based approach assumes a gradual weakening of the unstressed syllable nucleus due to overlap of adjacent consonantal gestures. According to the Gestural Score Model (Browman & Goldstein, 1990), gestures are performed by individual articulatory subsystems. Depending on the rate of speech, the model makes two different kinds of predictions: in faster or more casual speech, articulatory gestures can overlap to a greater or lesser extend (Munhall & Löfqvist, 1992). In theory, the second prediction is that in a slower rate of speech, the gestures for adjacent consonants in a cluster can become separated during the transition. Following Articulatory Phonology (AP), we hypothesize that depending on the degree of separation, gradually, transitional vowel traces can appear where they are not lexically specified. These vocalic traces are not controlled in comparison to the lexically specified schwas. Unlike AP, we further assume that 1.) not only the timing between articulatory gestures, but also tongue positioning (in particular, tongue lowering) is relevant for the production of a schwa. For instance, even if the velar and the tongue tip gesture in /gl/ are separated, schwa can only be present in the acoustics when the tongue back is lowered before /l/ closure. Without a tongue lowering the vocal tract should be too constricted and schwa will not realized acoustically. However, 2.) the perception of schwa may not necessarily depend on the realization of a schwa, since Price (1980) had suggested that lengthening of a liquid can also result in the percept of a syllable peak, thus vocalic traces must not be present.

To summarize, a slow speech rate can cause transitional schwa insertions between consonant clusters which appear as vocalic traces in the acoustic signal. Under such conditions, schwa falls out automatically without any specific articulatory target. On the contrary, articulatory schwa targets may be found in lexically specified schwa produced with a slow speech rate. Listeners may not only identify schwas on the basis of its occurrence in the acoustic, but also when listening to a lengthened liquid.

The present paper focuses on the temporal and positional articulatory characteristics and its corres-

ponding acoustics in the production of lexically specified and unspecified schwa in different speech rate conditions.

2 Our previous investigations

2.1 Evidence from acoustics

In a previous study (Jannedy, 1994) a corpus was read containing three pairs of the form CC or CəC whereby each member of the pair occurred within an identical segmental context. Six native speakers of a northern German dialect as spoken in the south of Hamburg, read the corpus each ten times in self selected speech rates. Speakers were instructed to produce rendition one and six at a normal rate, 2 through 5 increasingly faster relative to the previous reading and 7 through 10 slower and slower relative to the preceding reading. Duration measurements of each segment in the target words revealed a weakening of the unstressed syllable nucleus in faster rates of speech which can eventually result in the neutralization between the pairs of words only differing with regard to the schwa. In slower speech however, in some instances, we found the opposite effect, that is the appearance of vocalic traces between the two consonants in the onset cluster like [bər] instead of /br/. Based on the acoustic measurements, we hypothesize that a gestural reorganization based approach can best account for both of these rate effects found in German.

2.2 Evidence from perception

To test if this transition vowel can perceptually cause a confusion between the members of these "minimal pairs", Jannedy (1994) additionally carried out a forced choice perception test with 25 listeners whereby each stimulus was played twice and the listeners were requested to circle which word they heard. Identification ratings were generally good, however, there was some confusion on those tokens containing the consonant cluster but that were realized with a transitional schwa in slower rates of speech.

3 An articulatory study

To test the hypotheses that transitional (uncontrolled) schwas can fall out from imperfect gestural orchestration and positional characteristics (tongue lowering), we recorded articulatory data for 6 German speakers from a Northern variety of German by means of Electromagnetic Articulography (Carstens Medizinelektronik, AG 100). The speech material consisted of target words embedded in a carrier phrase with /bl/ vs. /bəl/ <bleiben> vs. <beleiben> and /gl/ vs. /gəl/ <gleiten> vs. <geleiten> in pre-tonic position, that is, before the main lexical stress of a word, and /bəl/ <Nabel>, /gəl/ <Nagel>, /dəl/ <Nadel>, and /n/ vs. /nn̩/ <kann> vs. <Kannen> in post-tonic position. Speakers realized all sentences in 3 self-selected speech rates, decreasing speed from normal to very slow. All tokens were repeated 7 times (9 words * 3 speech rates * 7 repetitions * 6 speakers = 1134 items). Three coils were attached to the tongue (tongue tip, tongue dorsum, tongue back), one to the upper lip, one to the lower lip, one to the lower incisors. Two coils served as references (one at the bridge of the nose and one at the upper incisors) to compensate for helmet movements. So far we will concentrate on 3 speakers' data.

3.1 Labeling procedures

The acoustic data were segmented and annotated using Praat 5.034. The pre-tonic clusters were labelled from the burst of C1 to the beginning of the second formant (F2) in the diphthong. If a schwa occurred in between, its on- and offset has been defined on the basis of the acoustic envelope and changes in formant structures. The post-tonic clusters were labelled similarly, but the beginning was defined as the closure onset. Additionally, the on- and offsets of all phones in the word were segmented. Several articulatory measures have been taken using Matlab. We will here focus on two temporal landmarks and their corresponding positions in /g(ə)l/ clusters (see Figure 1 below): the oral release of the tongue back sensor in /g/ (black arrow) and the time point where the tongue tip is first touching the palate in /l/ (red arrow).
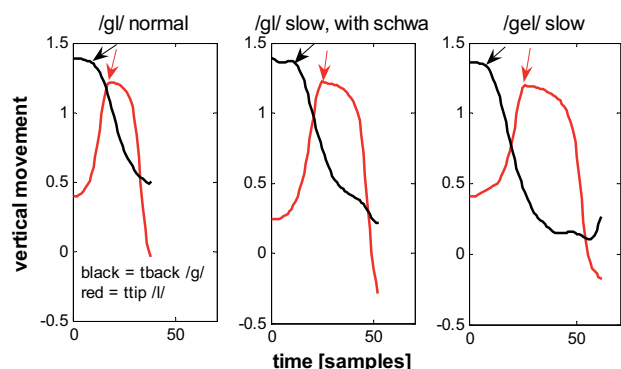
Figure 1. *Vertical tongue movement in cm for the tongue back coil (black) and the tongue tip coil (red). Arrows correspond to the 2 selected articulatory time landmarks.*

3.2 Overview of schwa realizations

Table 1 displays the number of actual vocalic traces found in the acoustics of each of the target words. So far schwa has been realized acoustically in most target words where it is lexically specified. However, it also occurred in the slow renditions of [gl]. Note that speaker sk shows the least variability in speech rate.

Table 1. *Number of vocalic traces found in the acoustic signal for each target word & speaker.*

Target word	Speaker vh	Speaker sk	Speaker ks
Pre-tonic			
[bl]-	1 of 21	0 of 21	4 of 21
[bəl]-	21 of 21	21 of 21	21 of 21
[gl]-	7 of 21	0 of 21	11 of 20
[gəl]-	21 of 21	21 of 21	17 of 17
Post-tonic			
-[n]	0 of 21	0 of 21	0 of 21
-[nən]	21 of 21	20 of 21	25 of 25
-[bəl]	19 of 21	20 of 21	19 of 21
-[dəl]	11 of 21	11 of 21	19 of 21
-[gəl]	16 of 21	19 of 21	19 of 21

3.3 Temporal results: Evidence for schwa insertion due to gestural separation?

In Figure 2 the duration of the time interval between velar oral release and /l/ target is displayed. Schwa traces found in /gl/ and /gəl/ are marked in orange

while realizations without schwa in /gl/ are plotted in black. A clear categorical boundary can be seen for the /gl/ data of all speakers occurring between 50 and 60 ms (longer durations showed consistently a transitional schwa –orange).
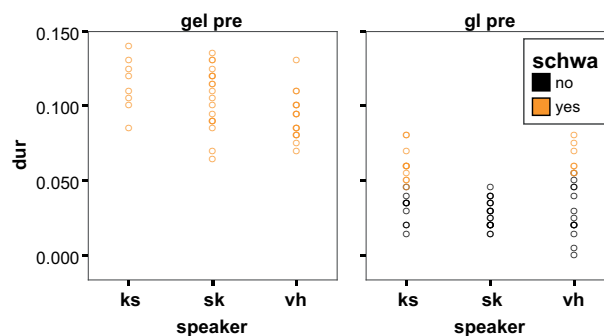

Figure 2. *Examples for the duration between velar release and /l/ target for different speakers in /gəl/ and /gl/. Orange dots = tokens with acoustic schwa, black = tokens without acoustic schwa; Pre = pre-tonic position.*

Based on these findings we conclude that the duration between adjacent consonant gestures in a cluster play an important role for a full acoustic realization of schwa.

However, this only holds true for the pre-tonic position. Figure 3 shows the results for the post-tonic position, in which schwa is always lexically specified.
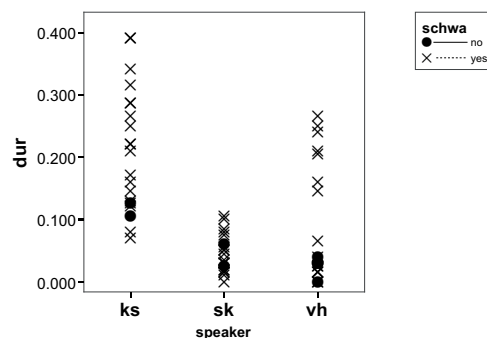

Figure 3. *Realizations in /gəl/with (x) and without (filled dots) schwa*

In this position schwa is found acoustically when the duration between the /g/ and /l/ gesture is very short and tokens including schwa may have a similar duration as tokens without schwa. From this perspective temporal constraints are not the only ones which lead to the realization of schwa.

3.4 Articulatory results

In order to discuss spatial articulatory characteristics with respect to schwa insertions, we subtracted the horizontal and vertical tongue positions of each coil at the /l/ target from their equivalent at the velar release. We expect that tongue back lowering after oral release is required for the realization of an acoustic schwa. Such tongue back lowering may be controlled in /gəl/, but falls out from the gestural separation in /gl/ with a transitional schwa.
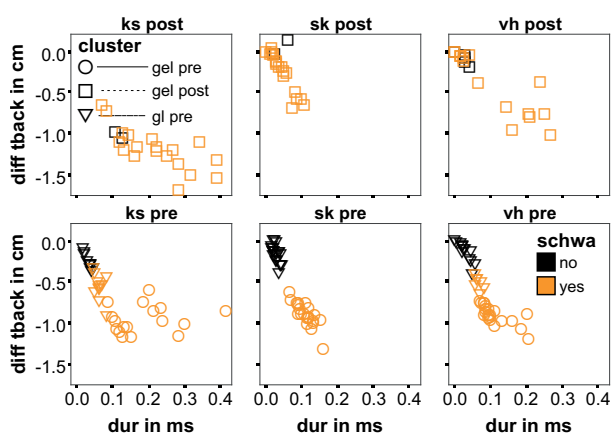

Figure 4. *Scatterplot with vertical tongue back movement (y-axis) plotted against the duration between oral release and /l/ target. Upper tracks: post-tonic position, lower tracks: pre-tonic position. Different speakers in columns. Black markers: schwa not realized in the acoustics, orange: schwa realized.*

Figure 4 displays the relation between tongue back downward movement and the duration between oral release and /l/ target. It can be seen that the tongue back moves least for /gl/ and most for /gəl/ (controlled schwa) with an intermediate step for /gl/ with transitional schwas. A similar relation has also been found for tongue back movements in the horizontal direction. So far it is unclear whether tongue lowering is an active strategy to produce schwa or whether it is simply the result of the fact that the tongue back is free to move towards the open vowel after velar release (including closed vowels may shed some more light on this issue). Figure 5 exemplifies the differences between /gl/ (upper tracks) and /gəl/ (lower tracks) for speaker vh. The most obvious difference between them is the low position (black arrows) of the tongue back coil when lexical schwa was present. In contrast, when schwa is not present, the tongue tip simply moves up whereas the velar closure stays relatively stable (circled in black).
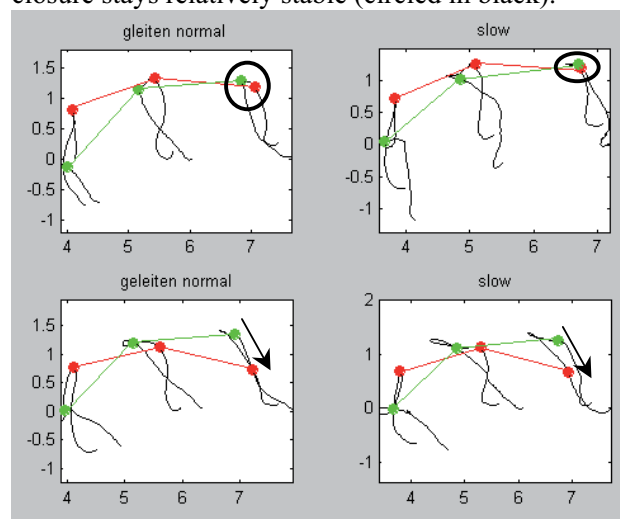

Figure 5: *Example for /gl/ without schwa (upper track) and /gəl/(lower track) in normal and slow speech condition; x-axis = horizontal movement, y-axis = vertical movement; dots: tongue target positions (tip, dorsum, back from left to right); green: /g/ target, red = /l/ target*

References

[1] Kloeke, W. v. L. *Deutsche Phonology und Morphology. Merkmale und Markiertheit.* Ling. Arbeiten 117, Max Niemeyer Verlag, Tuebingen, 1982.

[2] C. P. Browman, L. Goldstein. Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M.E. Beckman (eds.) *Papers in Lab. Phon. I: Between the Grammar and Physics of Speech*, 341-376. Camb. Uni. Press, Cambridge, 1991.

[3] K. J. Kohler. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. Hardcastle & A. Marchal (eds.) *Speech Production and Speech Modelling,* 69-92. Kluwer, Amsterdam,1990.

[4] K. Munhall, & A. Löfqvist. Gestural aggregation in speech: Laryngeal gestures. *Journal of Phonetics,* 20:111-126 1992.

[5] P. J. Price. Sonority and syllabicity: Acoustic correlates of perception, *Phonetica*, 37: 327-343, 1980.