OSU Working Papers in Linguistics 44 (105-124)

Rate Effects on German Unstressed Syllables*

Stefanie Jannedy
jannedy@ling.ohio-state.edu

Abstract: German is characterized by the rhythmic alternation of strong and weak syllables. Weak syllables contain short or reduced vowels like schwa. In some instances, the unstressed weak syllable nucleus can be the only difference between words that underlyingly contain a consonant cluster. Examples in German are *Kannen* 'cans, pitchers' contrasting with *kann* 'can (V)' or *beraten* 'to advise' contrasting with *braten* 'to fry'. In some instances, in 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. In slower speech, one might find an opposite effect, that is the appearance of vocalic traces between the members of an underlying consonant cluster. This transition vowel can perceptually cause a confusion in these "minimal pairs". Based on acoustic measurements, I will argue that gestural reorganization can best account for both of these rate effects found in German.

1. Introduction

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, 1989, 1990a). A gestural reorganization account assumes a gradual weakening of the unstressed syllable nucleus due to overlap of adjacent consonantal gestures. According to the Gestural Score Model (Browman & Goldstein, 1989, 1990a), 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 extent.

* I am thankful to my advisor, Mary Beckman, for discussions, valuable comments and more. Keith Johnson and Sun-Ah Jun also provided much helpful advice and have taken much time to discuss issues that arose while writing this paper. Beth Hume and Julie Boland provided helpful comments on an earlier draft of this paper. Jennifer Venditti, K. Brettonel Cohen, Chip Gerfen and Bettina Migge never got tired talking about reduction processes. I also need to thank my subjects (especially Ben) who repeated the corpus over and over or patiently listened to all the stimuli. The Department of Linguistics, Ohio State University and the University of Hamburg (Germany) are thanked for their support in this research. An earlier version of this paper was presented at the 68th Annual Meeting of the Linguistic Society of America, January, 1994.

In the case of a neutralization of a contrast, a gesture, in this case the one for the unstressed vowel, is completely overlapped and therefore hidden, so that no acoustic output is generated. In theory, a second prediction is that in a slower rate of speech, the gestures for adjacent consonants in a cluster can become separated during the transition. Depending on the degree of separation, gradually, vowel-like traces in the formant structure or even vowels of more than 20 ms in duration can appear where underlyingly not present. Phonological accounts on the other hand describe this phenomenon in terms of a categorical insertion rule (Hall, 1992).

Browman & Goldstein (1990a) provide x-ray microbeam data in support of their Gestural Score Model. In their example of the phrase *perfect memory,* the individual articulatory movements that were traced over time show that the closing gesture for the [k] in *perfect* [pʰɚfəkt] hides the closing part of the gesture of the [t], and the closure for the gesture of the bilabial [m] in the word *memory* [mɛmɚɹi] hides the release of the [t] on the tongue tip tier. There is no acoustic output from the alveolar gesture since the gesture of the [t] is hidden by the adjacent consonantal gestures on different, independent articulatory tiers. Similarly, Munhall and Loefquist (1992) provide data that suggests gestural overlap of adjacent glottal gestures in English. They had speakers say the phrase *kiss Ted* in various speech rates and focused on the glottal aperture at the word boundary between the [s] of [kɪs] and the aspirated [t] of [tʰɛd]. In the slowest renditions, they found two distinct glottal opening gestures, in an intermediate tempo, the gestures begin to blend and the one for the [s] becomes a shoulder of the gesture for the aspirated [tʰ]. In the fastest tempo, the two gestures have completely blended, so that only one glottal opening gesture is observable, and the [t] acoustically has lost the aspiration because no pressure could build up. Munhall's and Loefquist's data provide evidence for the blending of gestures on identical tiers.

With respect to the German weakening phenomena, the Gestural Score Model predicts gradually decreasing vocalic durations due to various degrees of overlap, or gradually appearing and increasing vocalic durations due to the gradual separation of articulatory gestures. Phonological accounts predict either a categorical deletion or a categorical insertion. In perception, we expect gradually poorer identification scores in faster rates for the word that contains the unstressed vowel since it becomes more and more reduced. The same prediction holds in cases where vowels gradually appear where not part of the underlying gestural score. If categorical phonological rules are at work, identification will be perfect if the vowel is present in the word that underlyingly contains the vowel or if the vowel is not present in the word that underlyingly does not contain the vowel. According to the Gestural Score Model, in the case of deletion, identification of words that only contrast by the appearence of the unstressed vowel should be impossible since there will be no contrast between forms that underlyingly do not contain a vowel and the forms that underwent vowel deletion. If there is a categorical insertion, we would expect identification to be impossible too, since there would not be a contrast between forms that underlyingly contained a vowel and the ones that underwent vowel insertion. To test these predictions, an acoustic study was performed.

1. Corpus and Methods

A paragraph was constructed that contained three target minimal pairs:

1.	Kannen	[k'anən]	'cans, pitchers'
	kann	[kan]	'can, (V)'
2.	geleiten	[gə.l'aɪ.tən]	'to accompany'
	gleiten	[gl'aɪ.tən]	'to slide'
3.	beraten	[bə.ʁ'a.tən]	'to advise'
	braten	[bʁ'a.tən]	'to fry'

Both members of each 'minimal pair' occurred within a context where the adjacent segments were identical. Six native speakers of a northern German dialect (as spoken in the south of Hamburg) read the corpus ten times each 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 the target words were done twice by the same person on a Kay Sonograph Spectrogram 5500-1. The measured values differed from the cross-checked values only minimally. Waveform and amplitude traces were in one display window and a wide-band spectrogram was displayed in a second window.

From each of the target words, the following measurements were taken: 1. total duration as defined from the release of the initial burst to the end of the final nasal. 2. the duration of the nasal sequence in *Kannen* and *kann*, and the duration from the onset of the second vowel to the end of the nasal in pairs two and three. 3. the VOT, from the release of the stop burst to the onset of voicing of the vowel, was taken for *Kannen* and *kann*. 4. duration from the initial burst to the end of the [a] in the first pair of words and to the end of the liquid [gl] in the second. For the *beraten* and *braten*, segmentation proved to be difficult and was cross-checked a third time with a computerized speech analysis system (Milinkovic). Previously, using the Kay, the end of the uvular fricative was determined by the dip in the amplitude trace. In C-speech, however, the end of the uvular fricative was determined by the onset of the decreased F2-bandwidth for the following [a]. The latter measurements were used in the plots of the production graphs. The appearance of a vowel in *braten* was judged by decreased F2 bandwidth right after the release of the initial voiced stop burst that then increases for the fricative. If this initial period in which we can observe an increased bandwidth was sustained for 20 ms or longer, it was judged to be containing a vowel.

For the perception test, 360 target words were spliced out of context, digitized, randomized and played back onto tape. There were also 120 filler words mixed into the randomized list of target words. Two tapes were prepared, each containing the same items but in a different order. 24 equally long blocks of 12 stimuli were recorded, each stimulus was played twice with an inter stimulus interval of two seconds. 25 native speakers of various German dialects participated in the forced choice identification perception test.

3. Results

The results of the production for *Kannen* and *kann* are shown in Fig.1. The durations of /ka/ of the monosyllable *kann* 'can, (V)' or of the disyllable *Kannen* 'cans, pitchers', are plotted on the x-axis as a measure of the speech rate. The target segments, that is for *kann* the /n/ (hollow circles) and for *Kannen* the nasal sequence /nən/ (filled circles) are plotted on the y-axis. The regression function indicates that as the rate of speech becomes faster and faster, the duration of the nasal sequence of *kann* also gradually shortens, but not as much as the one for *Kannen*.

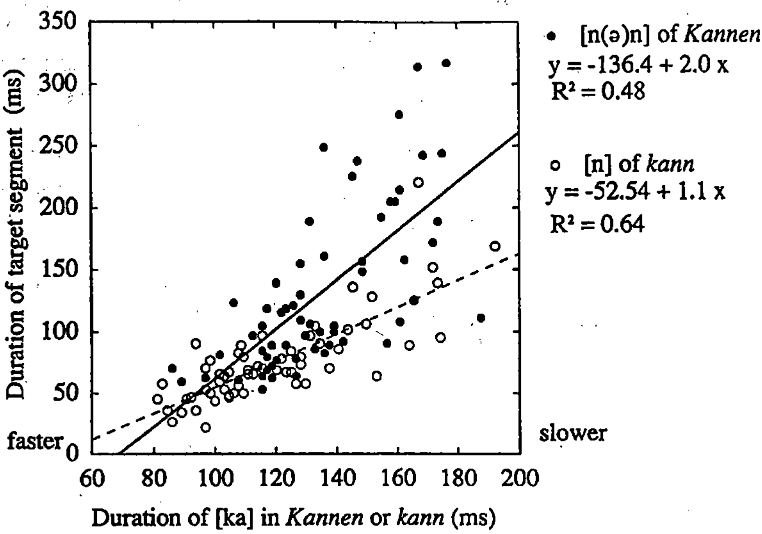

Fig.1: The duration of the nasal sequence of *Kannen* and *kann* is plotted against the remainder [ka] of *Kannen* or *kann* respectively.

The two regression functions for *Kannen* and *kann* cross as the rate of speech becomes faster. They show that the values for the duration of the [nən] and [n] incline toward each other, indicating that the duration of the nasal sequences in the monosyllable and the disyllable are of similar or equal duration in the faster rates but quite distinct in slower renditions. Yet, there is no categorical shift from presence to absence of the vowel in the nasal sequence /nən/ of *Kannen* since we do not find two clearly separated clouds of data. There were in total seven vocalic appearances of at least 20 ms in duration in all the tokens of *Kannen*. The spectrograms in Fig.2 show tokens of *Kannen* from continuous speech rates, uttered by the same female speaker. The token on the left was produced in rate 9 (slow rate), the one in the middle in rate 8 and the one on the right in rate 7, slightly slower than normal. Whereas there are very clear vocalic traces in the spectrogram on the left, the vowel is already shorter in the middle display and eventually totally disappears, as in the spectrogram to the right. The underlying vowel only appears in relatively slow and carefully articulated speech of two speakers in this study. However, the production data does not show any discontinuities but rather a very gradual shortening, as seen in the figure.

A paired t-test of the durations of the nasal sequences in *Kannen* vs. *kann* showed that the means of the samples were significantly different at the $p \leq 0.01$ level ($t = -8.56$, $p < .001$). However, the patterning of the individual datapoints supports the notion that the disappearance of the vowel, or the shortening of the [n]-sequence in *Kannen* is a gradual process rather than a categorical one.

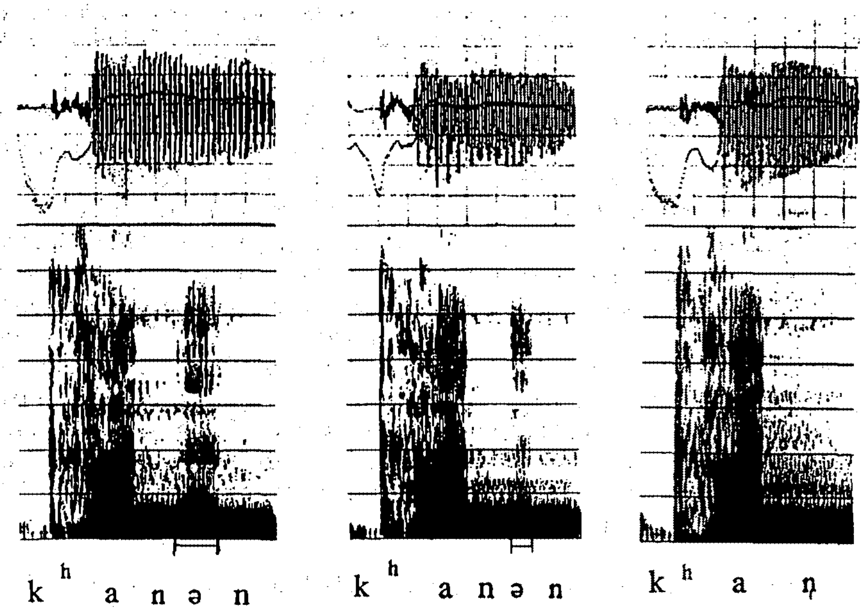

Fig.2: Spectrograms and waveform of tokens of *Kannen*, as produced by a female speaker in the rates 9, 8 and 7 from left to right.

The perception data in Fig.3 is based on the judgments of 25 listeners. The duration of the target sequences [n] and [nən] are plotted on the x-axis whereas the percent *kann* responses are plotted on the y-axis.

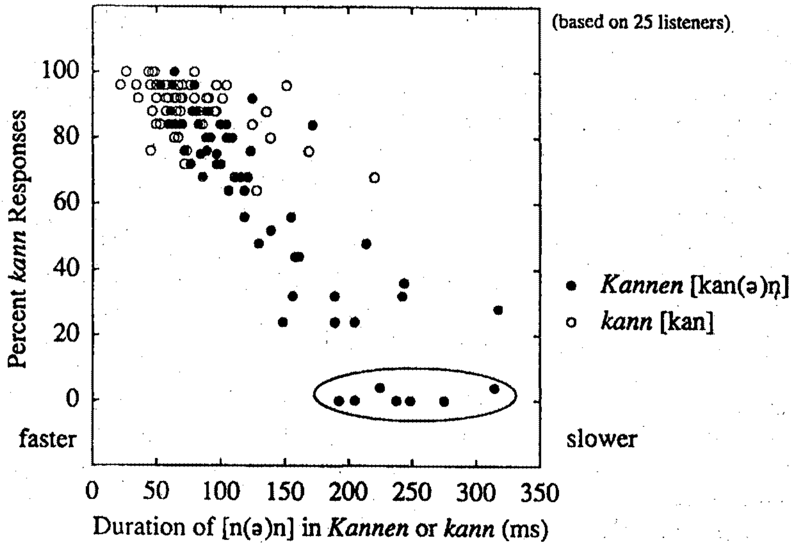

Fig.3: Perception results for *Kannen* (filled circles) and *kann* (hollow circles) as function of percent-correct of *kann*, showing how often *kann* was correctly perceived and how often *Kannen* was incorrectly perceived as *kann*. The tokens marked by the ellipse contain vowels.

The hollow circles represent tokens of *kann*, the filled circles stand for tokens of *Kannen*. The duration of the target sequence is plotted as a function of the percent correct responses for *kann*. Therefore, the patterning of the filled circles shows that as the rate of speech becomes faster, listeners increasingly incorrectly identify *Kannen* as *kann*. Only in the cases where the underlying vowel actually surfaces, the majority of listeners are able to identify these instances correctly. The tokens of *Kannen* containing the vowels are marked by the ellipsis. When the longer duration of the nasal indicates a second syllabic peak and takes on the function of the vowel, the listener's judgements tend to be rather incoherent. Just as in the production, we find overlap in the perception in faster rates. The correct judgments for *kann* always lie above 60%. The majority of judgments were made for *kann*, that is, listeners more often judged to have heard the monosyllable rather than the disyllable. Towards slower rates of speech, identification becomes slightly worse and values are scattered.

Fig.4 shows the production data for *geleiten* (filled squares) and *gleiten* (hollow squares). In both cases the target sequence is plotted as a function of the speech rate, manifested by the remainder of the word, which is [aɪtən]. The values are non-overlapping and throughout all speech rates, a difference between the two target portions of *geleiten* and *gleiten* is retained. A paired t-test shows that [gəl] and [gl] are significantly different at the .01 level ($t = 1.51$, $p = 0.0$).

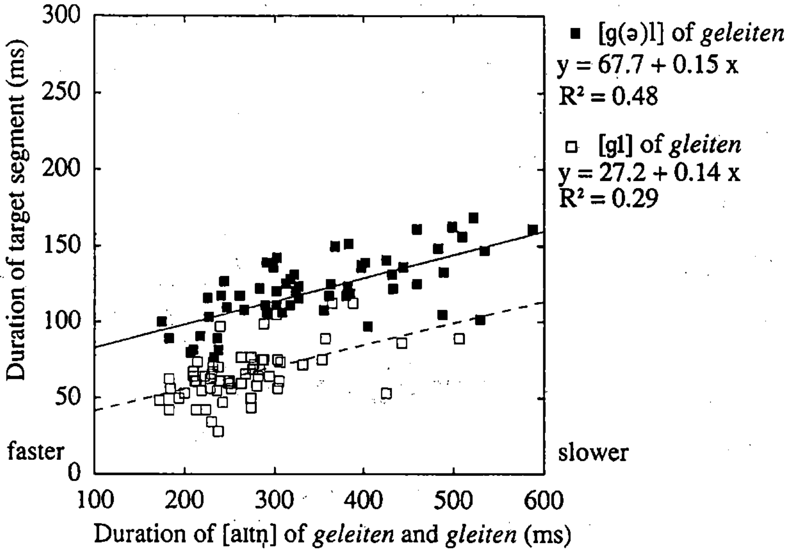

Fig.4: **The duration of the remainder is plotted against the target segments. Filled squares show cases of *geleiten* and hollow squares represent instances of *gleiten*.**

The regression functions indicate that as the rate of speech becomes faster, the values for the duration of the target sequences do not incline toward each other. All tokens of *geleiten* contained a vowel in all rates. If there was a categorical deletion, we would expect two distinct clouds of data for *geleiten* in production. This however is not the case; there is only one cloud of data for the underlyingly three syllable word. Also, the values for *gleiten* versus *geleiten* are fairly distinct, predicting good identification in the perception test. We can observe, however that as the rate of speech becomes faster, the duration of the target sequences [gəl] and [gl] becomes shorter in both cases, indicating that the faster rate of speech influences the duration of the target sequences.

Fig.5 shows the perception data for *geleiten* and *gleiten*. The duration of the target segments [gəl] (filled squares) and [gl] (hollow squares) is plotted as a function of the percent *gleiten* responses. Although identification is generally very good and we find two very distinct clouds of data which shows that the perception in these cases is rather categorical, there are a few cases of *geleiten* as well as *gleiten* that had a tendency to be misjudged more often. Apart from these few tokens, the majority of the words were correctly identified well above chance level. It is not clear what led listeners to make the few incorrect judgements.

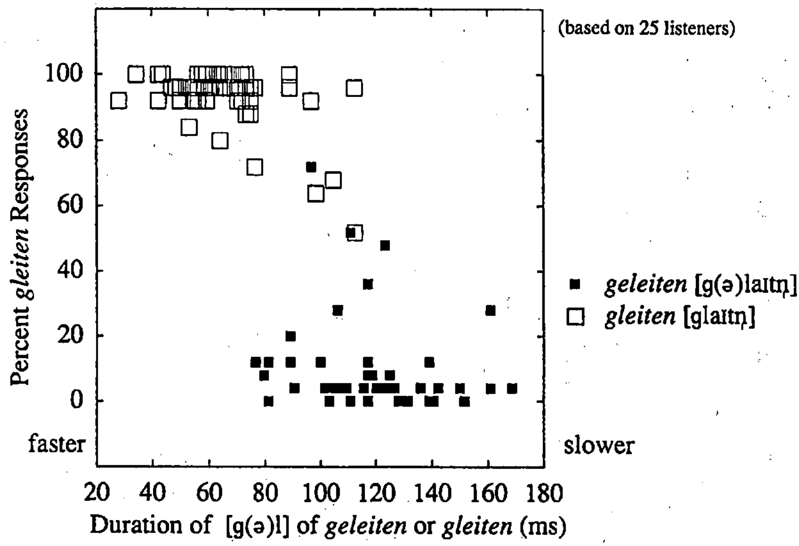

Fig.5: Perception data for *geleiten/gleiten*. The duration of the target sequence is plotted against the percent *gleiten* responses.

Fig.6 shows the production data for *beraten* and *braten*, for the six speakers. The graphs are presented individually to show the general tendencies for each speaker. These individual graphs show that speakers have different strategies and that one speaker's fastest rate can be another speaker's normal rate. The duration of the target segment [bəʁ] or [bʁ] is plotted on the y-axis and the remainder of the token, that is [atn̩] of either *beraten* or *braten*, is plotted on the x-axis. A paired t-test showed that these two target segments were significantly different at the .01 level ($t = 1.31$, $p < .001$). The filled triangles are tokens of the trisyllabic word *beraten* and the and hollow triangles symbolize tokens of *braten*. The filled circles stand for cases of *braten* where clear vowels of at least 20 ms of duration were found between the release of the voiced bilabial stop [b] and before the increase of the bandwidth of F2 during the uvular/velar fricative (due to decoupling of the front and the back cavity in the vocal tract). The values for the slopes and the intercepts of *beraten* and *braten* respectively are as follows for graphs one through six:

1. $y = 36.8 + 0.3x$, $R^2 = 0.53$; $y = 14.7 + 0.2x$, $R^2 = 0.56$;
2. $y = 82.4 + 0.1x$, $R^2 = 0.3$; $y = 52.2 + 0.1x$, $R^2 = 0.13$;
3. $y = 20.3 + 0.3x$, $R^2 = 0.71$; $y = 71.1 + 0.1x$, $R^2 = 0.4$;
4. $y = 128.2 + 0.1x$, $R^2 = 0.7$; $y = 27.8 + 0.2x$, $R^2 = 0.37$;
5. $y = -10.1 + 0.4x$, $R^2 = 0.72$; $y = 26.7 + 0.2x$, $R^2 = 0.46$;
6. $y = 22.0 + 0.4x$, $R^2 = 0.38$; $y = 38.8 + 0.1x$, $R^2 = 0.69$.

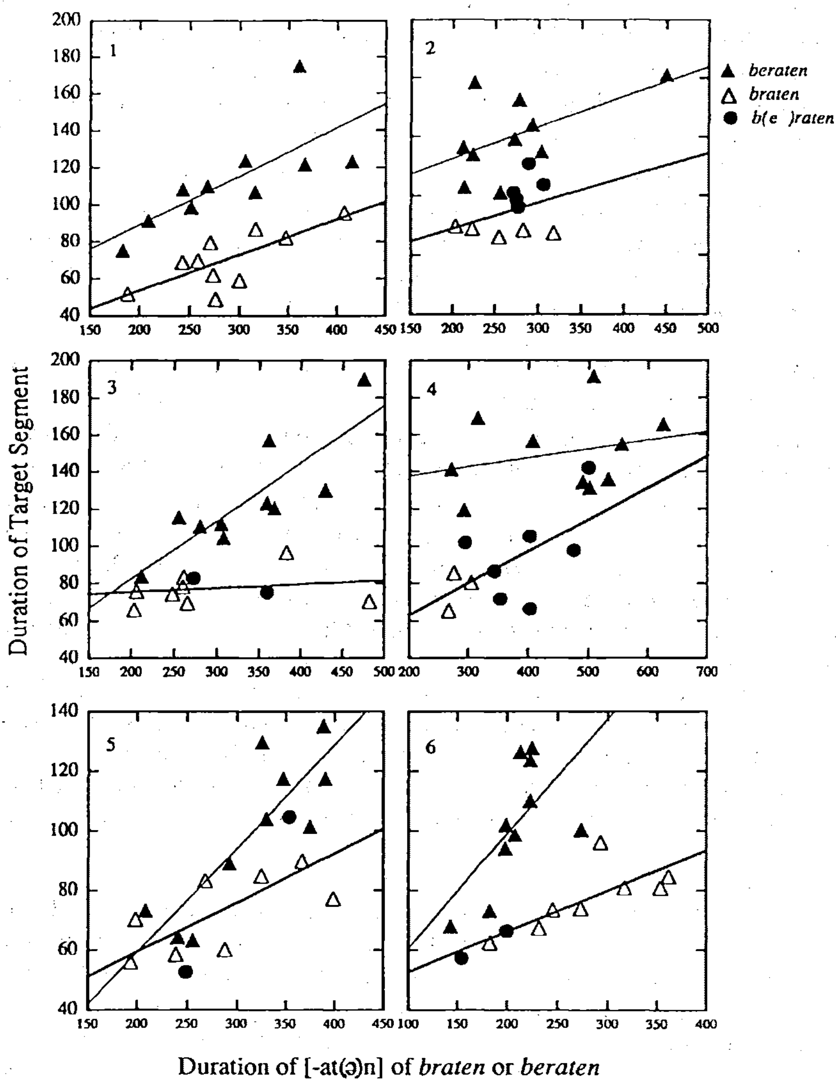

Fig.6: Production graphs for all six speakers for *beraten* (filled triangles) and *braten* (hollow triangles). The filled circles indicate the token where vowel-like segments appeared in the transition between [b] and [ʁ] in the *braten* cases. On the x-axis the duration of the respective remainder is displayed, and on the y-axis the duration of the target segment. The circles are included in the regression calculation.

There is not one consistent pattern for all six speakers in production. For speakers 3, 5, and 6, the regression functions incline towards each other in faster rates of speech, showing that as the rate of speech becomes slower, the values for the target segments [bəʁ] and [bʁ] diverge, with the values for [bəʁ] gradually increasing with a much steeper slope than the analogous values for [gl]. While for speakers 5 and 6 the duration of [bʁ] gradually increases as the rate of speech slows down, subject 3 shows only little variation in the duration of the target sequence throughout all speech rates. Speaker 4 shows that as the rate of speech slows down, especially the duration of [bʁ] gradually becomes prolonged, which is indicated by the steeper slope for [bʁ] than for [bəʁ]. For this speaker and also for speaker 2 there is much variation in duration of [bəʁ] in all rates of speech which is indicated by the scattering of data points around the regression line and the low R^2 values. Speaker 1 displays an almost constant increase in duration for both target segments as the rate of speech slows down.

Speaker 1 is the only subject who did not have any vocalic occurrences or even vocalic traces appear in the transition from [b] to [ʁ], in either rate of speech. Speakers 2, 3, 4, 5 and 6 on the other hand show some variability in the transition from [b] to [ʁ] in the underlying consonant cluster. Speakers 2 and 4 show the highest number of occurrences of transition vowels. There seems to be a strong tendency for these vowels to occur towards the slower rates of speech. Speakers 3 and 5 also show vocalic occurrences in the slower rates, however, there are also vowels found in faster rates. Speaker 6 displays the reversed pattern. This subject had vowels appearing only in the faster rates. Many more cases where somewhat dubious with regard to the set criteria since they showed a higher intensity but no increase in bandwidth of the F2 for the fricative. These cases are grouped with the ones that do not show the appearance of a vowel but only fricative striations (hollow triangles). In no case do we observe any abrupt discontinuities in duration as a function of the speech rate or as a function of the appearing vowel although there is an overall tendency for the transition vowel to occur in slower modes of speech.

The results of the perception data for *beraten* and *braten* are shown in the next graph. The percent-correct responses for *braten* are plotted as a function of the duration of the target segments [bəʁ] or [bʁ] respectively. The filled triangles represent instances of *beraten* whereas the hollow triangles represent instances of *braten*. The hollow circles represent those token of *braten* that contained a transition vowel. Although identification is generally good for both *beraten* and *braten*, more slower tokens of *braten* that contained a vowel were repeatedly misjudged as instances of *beraten*. This possibly indicates that listeners did not take speech rate differences into consideration.

However, there are also slower tokens of *braten* that did not contain a vowel and that were repeatedly misjudged. It remains unclear what cues listeners used to judge these tokens of *braten* as *beraten*. The identification rate for *braten* scatters between perfect (100%) and chance-level (50%). Two outlying tokens of *braten*, both containing a vowel, were often misjudged as *beraten*. However, there are cases where we find a transition vowel but that were identified very well as *braten.*

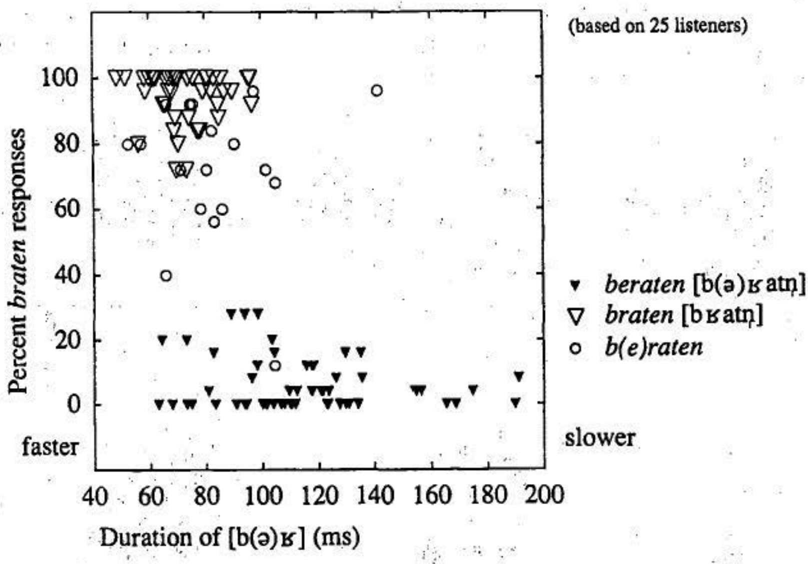

Fig.7: The duration of the target sequence is plotted against the percent *braten* responses. The hollow circles indicate which tokens of *braten* contained vowels.

Throughout all speech rates, identification of *beraten* was very good. *Beraten* was correctly perceived in at least 70% of the cases. There is a slight tendency for identification to get better the more the speech slows down, that is, the longer the target sequence becomes.

Fig.8 shows spectrograms of tokens of *gleiten* as well as *braten* to illustrate potential differences between tokens that were correctly perceived and those that listeners had trouble identifying correctly. All spectrograms in the upper panel were produced by the same male speaker. The spectrogram in the upper left panel shows an instance of *gleiten* (rate 5) that was correctly perceived by 24 out of 25 listeners. The one in the middle (rate 7) and on the right (rate 9) show two tokens of *gleiten* that were repeatedly misperceived as *geleiten* (in over 50% of the cases). Here, a sharp and sudden rise in intensity is reflected in the amplitude tracing. In the correctly perceived token, the amplitude rises into the intensity plateau of the following diphthong. In the mostly misperceived cases, the amplitude raises with a steeper slope and has already reached its maximum value during the transition into the vowel. Whether the difference in amplitude causes a difference in how the token is perceived is not clear at this point. Based on Price's (1980) experiments, however, amplitude was not a decisive factor. In the second row, we see three spectrograms of tokens of *braten* that were produced in different speech rates by the same female speaker. The spectrogram on the left shows a token produced in a faster rate than normal that was always correctly perceived. As expected, we do not find any vocalic segment between the word initial voiced bilabial stop [b] and the voiced uvular fricative [ʁ]. The spectrograms in the middle and to the right show two tokens that were frequently misjudged by listeners.

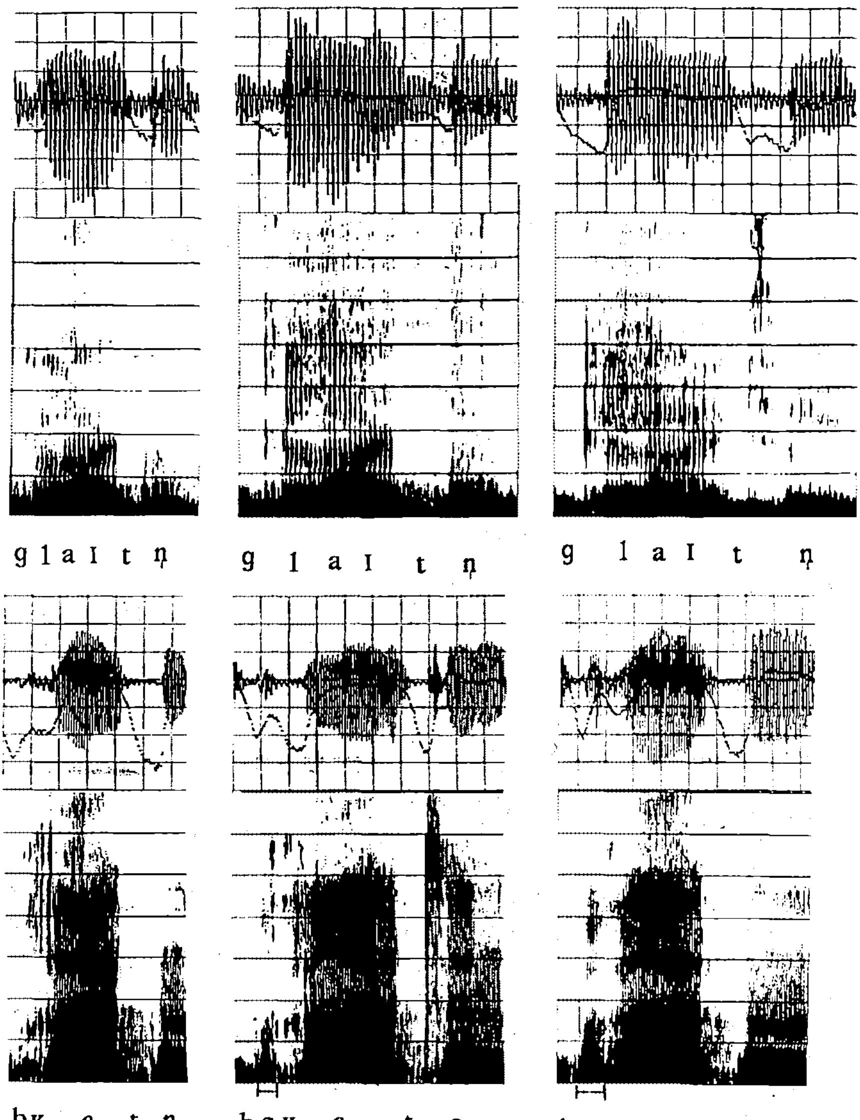

Fig. 8: Upper panel: Spectrograms of three tokens of *gleiten* (produced by the same male speaker). Left: rate 5; middle: rate 7; right: rate 9. Lower panel: Three tokens of *braten* (produced by the same female speaker). Left: rate 3; middle: rate 9; right: rate 8. In each panel, the tokens in the middle and to the right were misperceived most often.

The token in the middle was produced in a fairly slow rate (9) and the one on the right in rate (8). However, since the speakers were asked to read the corpus in self-selected speech rates, there is no absolute measure for the rate of speech. Therefore, rendition 8 can be slower than rendition 9 although it should be faster. In the second and third spectrogram, we do find vocalic traces between the initial cluster consonants. This finding is reflected in the waveform and the amplitude tracing as well. This transition vowel appears even stronger in the third display which was uttered slightly slower than the second one.

For a comparison, spectrograms for *geleiten* and for *beraten* (both in rate 1), produced by the same speakers as in the previous figure, are given in Fig.9. In the spectrogram on the left, we see clear vowel formants for the unstressed vowel, including a velar pinch, typical for a consonantal constriction in the velar region of the vocal tract. There is a sharp rise in the intensity level from the release burst of the [g] to a plateau at the unstressed vowel which is maintained throughout the [l] and the following diphthong.

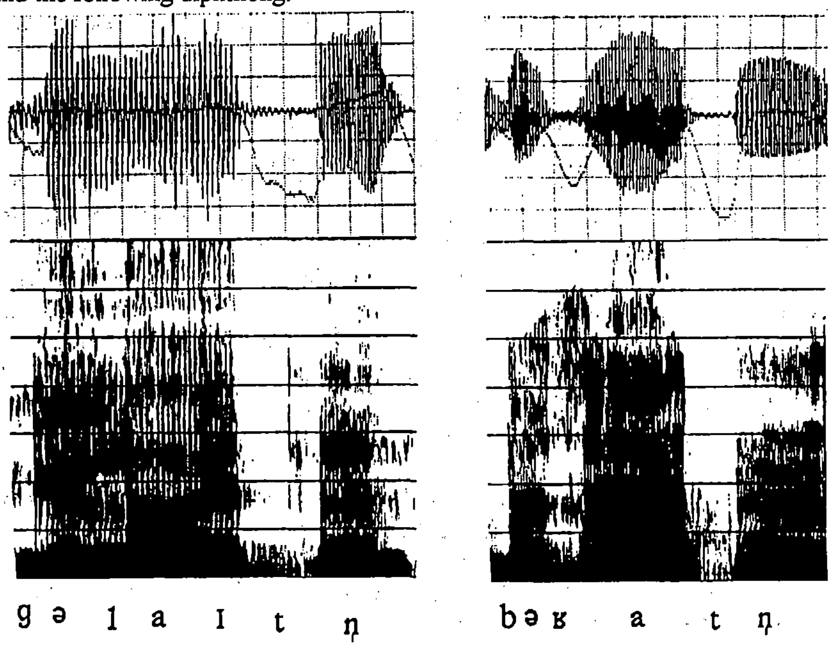

Fig.9: Spectrogram, waveform and amplitude tracings of *geleiten* (left) and *beraten* (right), produced in a normal rate by the same male and the same female speaker that produced the tokens in Fig.8.

On the right we find a clear vowel that is followed by frication with an energy concentration in the lower frequency area, typical for back fricatives. The waveform as well as the amplitude trace also clearly show the presence of the vowel before the voiced fricative. Compared to the spectrograms in Fig.8, the duration of the vowel is longer here. However, apparently, even the shorter duration of the vowel is sufficient to cause a perceptual confusion for the listeners. The seemingly longer duration of the [l] in *gleiten* in the misperceived tokens might have been interpreted as a syllabic peak. It remains unclear what parameters (amplitude, duration) listeners use. Even though the acoustic characteristics of *gleiten* vs. *geleiten* and *beraten* vs. *braten* show differences, the similarities are strong enough to cause perceptual confusion. To explore the specific perceptual cues listeners use is beyond the scope of this study.

4. Discussion

Even though Kohler (1992), in a critical commentary to the Gestural Score Model, argues that the mechanism of gestural overlap alone cannot explain connected speech processes as found in cases analogous to *Kannen*, the production data presented here for *Kannen* strongly favor a gestural overlap account of the observed reduction process over a categorical deletion process since no discontinuity is observed in production. Kohler (1992; 1990) gives examples like *kommen* 'to come' (/komen/ > [kommˌ]) and *Wagen* 'vehicle' (/vagen/ > [vagŋˌ]) for which he claims "a categorical change from progressive overlap [...] to reorganization" of gestures (1992:207). In his reorganization explanation he assumes that the apical closing gesture of the final [n] is deleted and that the opening of the velum is taking over the release of the stop. Browman & Goldstein on the other hand assume neither the addition nor a loss of gestures (see Johnson, Flemming and Wright, 1993:525). In the case of *Kannen,* as for the instances Kohler cites, it is still yet to be shown that phonetically there are no residual movements of the tongue apex left. According to the assumptions of the Gestural Score Model, gestures are abstract entities and the apical closing gesture remains present, however, somewhat reduced in magnitude (see Kingston, 1992). The burden of proof is on Kohler that the proposed theoretical gestural overlap account inaccurately explaines schwa reduction in this and analogous cases.

Fig.10 shows a hypothetical gestural overlap account, based on Browman & Goldstein's Gestural Score Model: as the rate of speech increases, the gestures for the alveolar nasals, produced on the tongue tip tier, as well as the gesture for the opening of the velum on the velic tier, move closer together, hiding the gesture for the unstressed vowel on the tongue body tier. In faster rates or casual speech, the gestures of the alveolar nasals and the velic gestures have completely overlapped the gesture for the unstressed vowel. The vocalic gesture is hidden behind the consonantal gestures and not deleted, as has been claimed in phonological accounts.

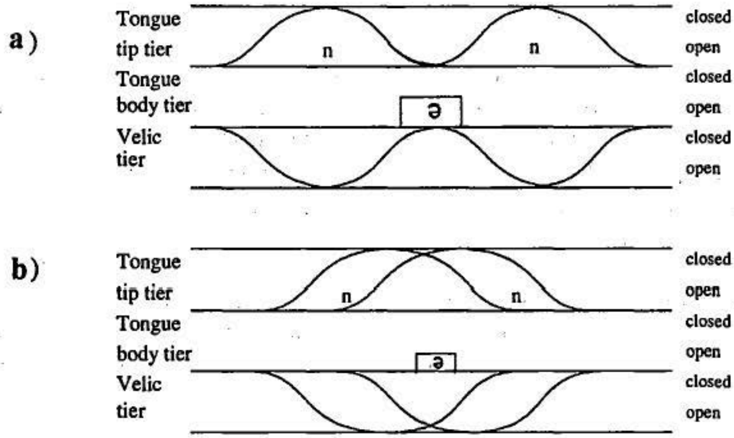

Fig.10: Hypothetical Gestural Score showing in a) two alveolar gestures on the tongue tip tier for the nasals and the unstressed vowel on the tongue body tier. In b) a situation where the gestures for the nasals overlap and hide the vocalic gesture, but where some durational difference in the nasal sequence vs a single nasal is still preserved.

Various examples cited in the literature for English and other languages (see Beckman, this volume) show cases where articulatory gestures are overlapped and become hidden (Browman & Goldstein, 1989, 1990a; Munhall & Loefquist, 1992; Jun & Beckman, 1993) so that their acoustic output is much reduced which then can potentially result in the total loss of the perceptability. Also, Browman & Goldstein (1990b:314/315) showed that the identification function in synthesized stimuli on a continuum of gestural overlap between the gestures for the [b] and the rhotic (American) [ɚ] from *bray* to *beret* (generated from calculations of 'the task dynamic and vocal tract models') shifted at 0 ms overlap (perfect alignment) for four out of six listeners. Price (1980) provides similar results by modeling gestural overlap with varying the duration of [l] in synthesized stimuli of *plight* and *polite*.

No cases cited in the literature give examples or show evidence for the gradual appearance of a vocalic segment due to gestural separation even though such cases are not ruled out by Browman & Goldstein (1990b:318; 1992b:53) e.g. as a possible source for sound change. The evidence from (especially slower renditions of) *braten* favor an account of a gradual separation of articulatory gestures that results in the gradual appearance of vowel-like traces in the formant structure or even a vowel. The acoustic output during the transitions from one articulatory gesture to the next one is then misperceived. In contrast to the [l] in *plight* (Price, 1980) or the [ɚ] in *bray* (Browman & Goldstein, 1990b) where sonorant liquids were lengthened to the point that the duration gave listeners an increased percept of sonority and hence a syllabic peak (both these sounds have the property of being able to become syllabic), the German voiced uvular fricative does not seem to have the property of becoming syllabic with an increased duration (see Kohler, 1991 on German [ʁ] vocalization).

An explanation for the appearance of the transition vowel in *braten* (in comparison to *gleiten*) comes from the aerodynamic properties of the voiced uvular fricative [ʁ]. Voiced fricatives have in general been found rather rarely in the world's languages (Ohala, 1983) since a fine tuning of the airstream mechanism is required to produce this class of sounds. Air flow must be high enough to cause frication at a given constriction within the vocal tract and at the same time, the constriction at the glottis needed for vocal fold vibration impedes airflow. These two articulatory goals are difficult to achieve simultaneously and therefore voiced fricatives display greater instability in production since only a slight change in one of the articulatory parameters can cause great variability and discontinuities in the acoustic output (Stevens, 1972).

The variability in production for *braten* can thus be explained as follows: right after the release of the bilabial occlusion for [b], there will be stronger airflow. After the air pressure has leveled out (within approximately 3 to 5 glottal pulses), the voicing sets on and the uvular constriction for the fricative can be formed again. In these cases where the articulatory gestures are misaligned, vowel-like formant structures are produced. Vowel durations between 30 and about 60 ms were found where actually no vowel is underlyingly present. In Fig.11, the appearance of this transition vowel is explained by means of Browman & Goldstein's Gestural Score Model: as the closure of the [b] on the lip tier is released, the gesture for the uvular fricative on the tongue body tier has already set on and (almost) come to a closure on the tongue body tier. There is no transition vowel produced here.

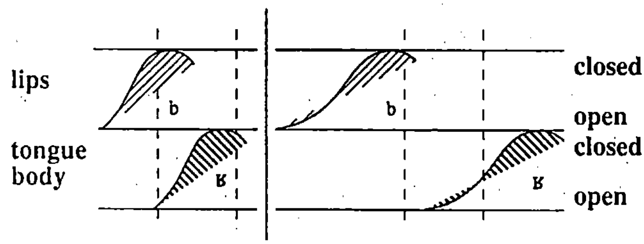

Fig.11: Hypothetical Gestural Score for the sequence [bʁ] in *braten*. On the left, the gestures are phased so that no vocalic output will be generated since the [ʁ] has already come to a closure. On the right, the gesture for [ʁ] only sets on after the gesture of the [b] has been released.

On the other hand, in the right panel, as the lips open after the bilabial closure for the [b], the tongue body has barely begun to rise, and the separation of gestures in time creates a transition vowel. Therefore, gestural separation appears to be responsible for generating an unintended vocalic segment. Even when a transition vowel appears, there will be no gesture for the vowel since it is not part of the gestural score for [bʁ].

This misalignment of gestures and the instability of the articulation of the voiced uvular fricative is not an aerodynamic artifact since for example speaker 1 did not show any cases where formant structures in the acoustic output were visible. Nor does it only occur in slower rates of speech since speaker 6 showed a reversed pattern. The gestural separation and the resulting transition vowels are a result of the phasing of the release of the bilabial occlusion and the onset of constriction and timing of the laryngeal gesture for [ʁ]. Speakers might have some intuitive knowledge based on their experience with language about how much variation is allowed in production to still match the output goal (see Hawkins, 1992:57). Even though clear vowels appeared in the acoustic signal, the listener can apparently to some degree compensate for this and still perceive these tokens as instances of *braten*. The listener might have taken the durational difference of the vowel into account for which he or she has some kind of expectation. However, this can only be hypothesized at this point. In no case of *beraten* the schwa was reduced to such an extent that the acoustic output was completely hidden. This might be explained along the same lines as the appearance of vocalic traces and vowels in *braten*.

During the repositioning of the articulators from one articulatory gesture to the next, the tongue moves through a "neutral position" (Browman & Goldstein, 1992a:55; also see Barry, 1992 for a discussion) and the transition vowel will be generated. Phonological accounts treat the appearance of a vowel as a categorical insertion (Hall, 1992). However, temporal or syntagmatic coordination (Browman & Goldstein, 1992b) and therefore, gestural separation and the misalignment of articulatory gestures can be modeled by the Gestural Score Model. It is crucial though that the transition vowel found in some cases of *braten* does not have a target (Browman & Goldstein, 1992a) since it is not part of the underlying form.

Based on the findings of gestural overlap in faster rates of speech as exemplified in the case of *Kannen*, the appearance of the vocalic traces in slower renditions of *braten* can be now explained without positing a categorical vowel insertion rule but by the reverse mechanism of gestural overlap, that is gestural separation. The Gestural Score Model therefore provides a unified account of gestural reorganization for the observations that are traditionally explained by processes of deletion (Kloeke, 1982) or insertion (Hall, 1992; Strauss, 1982).

According to Lindblom's (1990) 'Hyper- and Hypoarticulation' Theory, speech production varies along a continuum of hypo- and hyperarticulated speech (see also Johnson, Flemming and Wright, 1993). The speaker takes the communicative demands of the situation into account and strives for sufficient discriminability. This process is controlled by output oriented feedback mechanisms. Depending on the amout of effort a speaker puts into the production of speech, the articulatory target can be undershot. In this study, for *braten*, vowels were found not exclusively but predominantly in slower renditions of speech which will be equated with more careful, clear or hyperarticulated speech. In this study though, it was found that carefully produced speech (hyperarticulation) can result in the appearance of vowels which in turn can mislead the listener. This of course cannot be the goal for the speaker. However, the data presented for *braten* does not argue against the theory that phonetic targets are hyperarticulated. The misaligned transitions rather than an overshoot of the phonetic target, that is the syntagmatic relationship (Kohler, 1986) between one gesture's offset and an adjacent gesture's onset, cause the appearance of unintended vocalic traces or vowels which potentially cause perceptual confusion. Interestingly, the appearence of the vowel occurred in the case of [bʁ], the sequence with an aerodynamically unstable sound. Note also that the appearance of the vowel was not perfectly correlated with the rate of speech or the care of production. Previous studies did not specifically look at cases that display aerodynamic instabilities of the consonantal gestures involved forming the consonant cluster. To verify that this is a necessary prerequisite for non-underlying vowels to appear, more evidence is needed. It is conceivable also that the obstruent status of the [ʁ] influences the appearance of the transition vowel. This, however, needs more exploration.

5. Conclusion

Evidence for the applicability of the Gestural Score Model for German was provided. The production as well as the perception data presented for *Kannen* argue for a gradual reduction of the unstressed syllable nucleus rather than a categorical deletion process. Furthermore, for *braten*, in some instances unpredicted (i.e. non-underlying) vowels and vowel-like traces were produced. Here the opposite mechanism appears to be at play, that is a gradual separation of adjacent articulatory gestures due to a misalignment in temporal coordination. Therefore, it is concluded that the Gestural Score Model can account for both, the gradual reduction as well as the gradual appearance of vowels in German, whereas the phonological accounts needs to posit a categorical insertion as well as a categorical deletion rule.

6. References

Barry, W. (1992) Comment on "Targetless" schwa: an articulatory analysis. In G. J. Docherty & D. R. Ladd, eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody,* 65-67. Cambridge: Cambridge University Press.

Beckman, M. (this volume) When is a Syllable not a Syllable? OSU Working Papers in Linguistics, 44.

Browman, C. P., Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology*, 6, 201-251.

Browman, C. P., Goldstein, L. (1990a) Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman, eds., *Papers* *in Laboratory Phonology I: Between the Grammar and Physics of Speech*, 341-376. Cambridge: Cambridge University Press.

Browman, C. P., Goldstein, L. (1990b) Gestural specification using dynamically-defined structures. *Journal of Phonetics*, **18**, 299-320.

Browman, C. P., Goldstein, L. (1992a) "Targetless" schwa: an articulatory analysis. In G. J. Docherty & D. R. Ladd, eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, 26-56. Cambridge: Cambridge University Press.

Browman, C. P., Goldstein, L. (1992b) Articulatory Phonology: An Overview, *Phonetica*, **49**, 155-180.

Hall, T. A. (1992) Syllable Structure and Syllable-related Processes in German, *Linguistische Arbeiten 276*. Tuebingen: Max Niemeyer Verlag.

Hawkins, S. (1992) Comment on "Targetless" schwa: an articulatory analysis. In G. J. Docherty & D. R. Ladd, eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, 56-59. Cambridge: Cambridge University Press.

Johnson, K., Flemming, E. & Wright, R. (1993) The hyperspace effect: Phonetic targets are hyperarticulated, *Language*, **69**, 505-528.

Jun, S.-A., Beckman, M. E. (1993) A gestural-overlap analysis of vowel devoicing in Japanese and Korean. Paper presented at the 1993 Annual Meeting of the Linguistic Society of America, 7-10 January 1993, Los Angeles, CA, USA.

Kingston, J. (1992) Comment on "Targetless" schwa: an articulatory analysis. In G. J. Docherty & D. R. Ladd, eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, 60-65. Cambridge: Cambridge University Press.

Kloeke, W. v. L. (1982) Deutsche Phonology und Morphology. Merkmale und Markiertheit. *Linguistische Arbeiten 117*. Tuebingen: Max Niemeyer Verlag.

Kohler, K. J. (1986) Invariance and Variability in Speech Timing: From Utterance to Segment in German. In Perkell, J. S. & Klatt, D. H., eds., *Invariance and Variability in Speech Processes*, 268-289.

Kohler, K. J. (1990) Segmental Reduction in Connected Speech in German: Phonological Facts and phonetic Explanations. In W. J. Hardcastle & A. Marchal, eds., *Speech Production and Speech Modeling*, 69-92. Amsterdam: Kluwer.

Kohler, K. J. (1991) Synthesis of German /r/ in Text-to-Speech. In *Proceedings of the XIIth International Congress of the Phonetic Sciences*, **3**, 490-493. Aix-en-Provence, France, 1991

Kohler, K. J. (1992) Gestural Reorganization in Connected Speech: A Functional Viewpoint on 'Articulatory Phonology'. *Phonetica*, **49**, 205-211.

Lindblom, B. (1990) Explaining Phonetic Variation: A Sketch in the H&H Theory. In W. J. Hardcastle & A. Marchal, eds., *Speech Production and Speech Modeling*, 403-439. Amsterdam: Kluwer.

Munhall, K., Loefquist, A. (1992) Gestural aggregation in speech: laryngeal gestures. *Journal of Phonetics*, **20**, 111-126.

Ohala, J. J. (1983) The Origin of Sound Patterns in Vocal Tract Constraints. In P. F. MacNeilage, ed., *The Production of Speech*, ch. 9, 189-216.

Price, P. J. (1980) Sorority and Syllabicity: Acoustic Correlates of Perception, *Phonetica*, **37**, 327-343.
Stevens, K. N. (1972) The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data. In *Human Communication: A Unified View.* Chapter 3, 51-66.
Strauss, S. L. (1982) Lexicalist Phonology of English and German. Dordrecht: Foris.