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

ETRW on Sound Patterns of
Spontaneous Speech (SPoSS)
La Baume-les-Aix, France
September 24-26, 1998

HOMORGANIC AND QUASI-HOMORGANIC V_1 ## V_2 SEQUENCIES IN AUSTRIAN GERMAN

Sylvia Moosmüller
Acoustics Research Department, Austrian Academy of Sciences, Vienna, Austria
E-mail: moosm@kfs.oeaw.ac.at, Fax: +43 1 4277 9296

RESUME

La présente étude traite des divers pas progressifs intermédiaires des procès d'assimilation des séquences V_1 ## V_2 homorganiques et quasi-homorganiques. Il est entrepris l'expérience de subdiviser la gradualité de ce procès en des pas progressifs intermédiaire, commençent par la réalisation canonique [V_1?V_2] et arrivant à l'assimilation totale dans laquelle plus aucune limite entre les segments peut être observé.

1.INTRODUCTION

Canonically, V_1 ## V_2 sequences should be realized as [V_1?V_2], the glottal stop indicating the segment boundary. Speaking both in terms of production and perception, maximal contrast can be observed between the two adjacent vowels. As a function of various linguistc and socio-psychological variables ([1], [2]), reduction and assimilation processes can be observed ([3], [4], [5], [6]). However, the reduction of [V_1?V_2] does not follow exactly the same pattern as has been described for intervocalic stops ([5], [7]), as the glottal stop in a first step is replaced by creaky voice. Nevertheless, intervocalic sound energy difference [5] remains an important parameter in the process of gradual assimilation. The endpoint of the gradual sequence of assimilation processes is the loss of the glottal stop and the total assimilation of the two adjacent vowels, as can be seen in the following spectrogram:

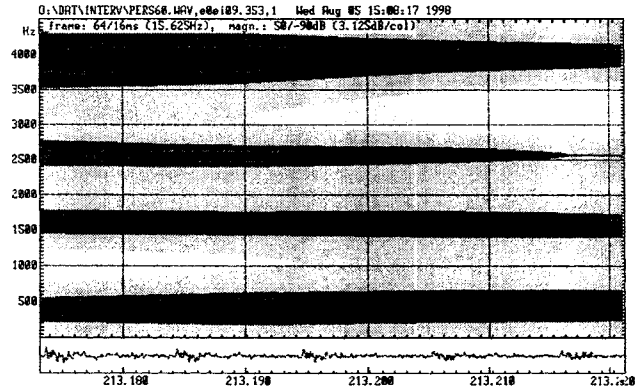

figure 1: spectrogram of the sequence /ɛ##aɛ/ taken from the sample "lade Leute ein" (invite people). x-axis: time, y-axis: frequency in Hz.

The spectrogram shows the sequence /ɛ##aɛ/ taken from the sample "lade Leute ein" (invite people). It has to be mentioned that in Austrian German diphthongs are usually monophthongized, especially in weak positions [8]. It can, therefore, be assumed that the diphthong /aɛ/ is realized as [ɛ], resulting – for the example under consideration – in a sequence of two homorganic vowels: "Leut[ɛ] ## [ɛ]n". As can easily be seen from the spectrogram (figure 1), no segment boundary is visible any more and, furthermore, the two vowels have assimilated completely.

The aim of the present paper is to show the diverse intermediate steps of assimilation processes, starting with the canonical realisation [V_1?V_2] with a concommitant clear segment boundary and ending with the total assimilation of the two adjacent vowels, where the segment boundary can no longer be determined. As a matter of course, degree and direction of assimilation obey the principles which have often been described in the literature, as e.g. strength of position: weak positions are more prone to assimilation than strong positions. Furthermore, in cases of non-homorganic V_1 ## V_2 sequences, the weak vowel assimilates towards the prosodically stronger one. Segment duration and strength of position are interdependent. Therefore, the duration of the discussed example (figure 1) is only 61ms, whereas the duration of a similar sequence, "ein[ɛ] ## [aɛ]nfache" (a simple), which has been realized with a glottal stop between the final [ɛ] and the initial [aɛ], amounts to 239 ms. All theses facts as well as extralinguistic factors influencing assimilation processes can be deduced from theory (e.g. Natural Phonology, cf. [1], [9], [10]) and won't be dealt with in the present study. In the discussion of the diverse degrees and modes of assimilation of V_1 ## V_2 sequences therefore, no reference will be made either to linguistic (segmental or prosodic) or to extralinguistic phenomena influencing the assimilation under consideration. However, an attempt will be made to set up – if possible – gradual steps on the vertical axis of no assimilation to total assimilation.

2.METHOD

The spontaneous speech (semi-structured interviews) of five male Austrian Standard speakers has been used [2]. The signal was digitized at 16 kHz, 16 Bit, by means of the Acoustic Workstation S_Tools [11]. The following V_1 ## V_2 sequences (250 in total) have been analyzed:

V_1 ## V_2	canonical transcription	example	translation
er ## a	ɐ?a	der andere	the other one
ar ## a	a?a	war also	was also

er ## e/ei	ɐʔɛ/aɛ	der einfach	the simple
ar ## e/ei	aʔɛ/aɛ	war eine	was a
e ## a	ɛʔa	nehme an	(I) assume
a ## e/ei	aʔɛ/aɛ	a Eierspeis	an omlette
e ## e/ei	ɛʔɛ/aɛ	gerne esse	like to eat
a ## a	aʔa	da alle	there all

The sequences have been analyzed according to the presence or absence of a glottal stop or creaky voice. The first three formant frequencies and amplitudes were calculated by means of LPC analysis, 22 coefficients and a pre-emphasis 0.9. Measurements were taken at the center of V_1, V_2 and at the (assumed) transition part. In order to evaluate the fine structure of V_1 ## V_2 transitions, typical tokens have been analyzed in detail. Frame by frame formant frequency amplitude contours are shown in figures 5 to 9. The frame width was 64ms and frame shift varied from 2ms to 8ms in order to obtain linear time scale standardization. In cases of transition of creaky to modal voice, fundamental frequency has also been measured.

3.RESULTS

3.1.Segment boundary marked by a glottal stop

From a phonological point of view, this realization displays maximal contrast: voice – voicelessness – voice, there is no doubt about the position of the segment boundary. Moreover, [V_1?V_2] represents the canaonical realization. Nevertheless, the glottal stop realization has been rarely observed in the analyzed speech material.

3.2.Segment boundary marked by creaky voice

As a next step, the state at which the glottis is kept closed – a state which lasts from about 40 ms to 80ms – is substituted by creaky voice which functions as a transition phase from V_1 to V_2. Usually the last glottal period of V_1 and the first glottal periods of V_2 are marked by irregular duration. The segment boundary therefore, becomes a bit blurred, especially in homorganic V_1 ## V_2 sequences.

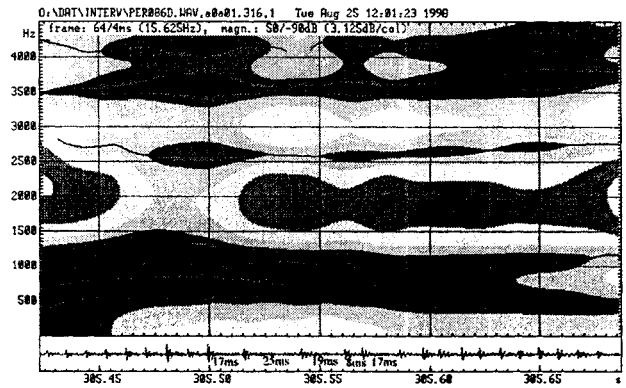

figure 2: spectrogram of the sequence /a##ɑɔ/ taken from the sample "Firma aus" (from the firm). x-axis: time, y-axis: frequency in Hz.

Figure 2 shows the spectrogram and the waveform of the sequence /a##ɑɔ/ taken from the sample "Firma aus" (from the firm). At the transition from V_1 to V_2 irregularities in the periodicity of the waveform and a drop in the amplitude of all formant frequencies can be observed. With respect to V_1, the drop in amplitude amounts to 2dB for F1, 1dB for F2 and 8dB for F3. The amplitude values increase again in V_2 (8dB for F1 and 7dB for F3), with the exception of the amplitude of F2, which is marked by a continuous drop. As far as vowel quality is concerned, the gradual movement from the front vowel [a] to the back vowel [ɑ] is still maintained, i.e. no assimilation of vowel quality has taken place. Therefore, the transition part is characterized by three parameters:

1. irregularity in the periodicity of the waveform with concommitant effect on fundamental frequency,
2. differences in formant frequency amplitude, affecting all formants, and
3. change in vowel quality.

Taking these three parameters into account, the segment boundary can be determined quite accurately.

However, as the airflow is no longer suddenly interrupted, assimilation with respect to vowel quality (i.e. formant frequencies) might be observed already at this stage.

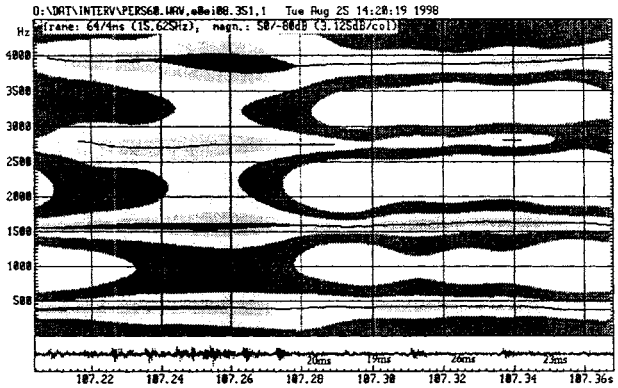

figure 3: spectrogram of the sequence /ɛ##aɛ/ taken from the sample "Leute einlad" (invite people). x-axis: time, y-axis: frequency in Hz.

Figure 3 shows the spectrogram and the waveform of the sequence /ɛ##aɛ/ taken from the sample "Leute einlad" (invite people). As can be seen from the formant frequency contour, no change in vowel quality takes place. However, V_2 is easily recognizable, as the whole sequence of V_2 is marked by creaky voice. Furthermore, a decrease of all formant frequency amplitudes of V_2 as compared to V_1 can be observed. This example does not exhibit a transition phase, V_2 starts at the beginning of creaky voice. Although only two parameters fit into the separation of V_1 and V_2, segment boundary can be determined more easily than in the example described in figure 2, because the transition part is missing. Nevertheless, compared with the example of figure 2, this sample can already be evaluated as a step further in the direction towards total assimilation.

3.3.Transition from creaky to modal phonation

Gradually, the irregularity of the periodicity of the waveform is substituted by modal phonation. In certain cases, two or three glottal periods might still be marked by a slightly longer duration with the concommitant slight decrease in fundamental frequency. Thus, creaky voice is substituted by a simple decrease in fundamental frequency which now functions as the transition phase.

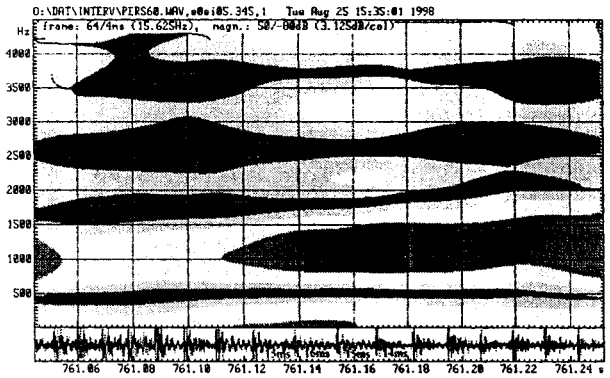

figure 4: spectrogram of the sequence /ɛ##aɛ/ taken from the sample "gute Einstellung" (good attitude). x-axis: time, y-axis: frequency in Hz.

Figure 4 shows the spectrogram and the waveform of the sequence /ɛ##aɛ/ taken from the sample "gute Einstellung" (good attitude). The transition phase is marked by a drop in fundamental frequency from 100Hz to ~ 65Hz. In V_2, F_0 rises to ~ 80Hz again. A movement especially of the second formant can also be observed. However, the first part of the diphthong /aɛ/ from "Einstellung" (attitude) has the same formant frequency values as the final /ɛ/ from "gute" (good). In this case, therefore, formant frequency contour does not help to determine the segment boundary. A further difference from the examples discussed above can be also observed: Both F_1 and F_2 are marked by a continuous drop in amplitude starting at the transition phase, a subsequent

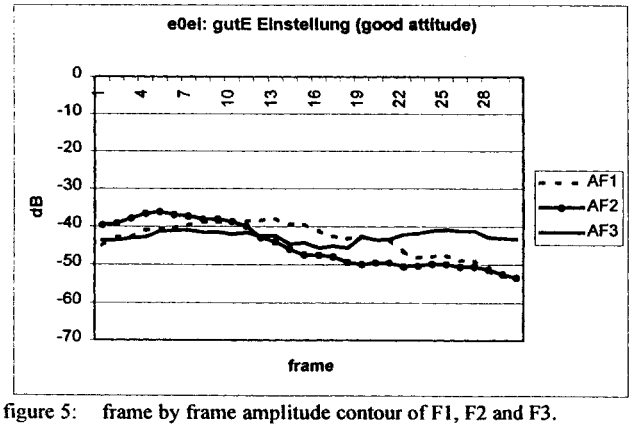

figure 5: frame by frame amplitude contour of F1, F2 and F3.

clime after the transition phase takes place within F_3 only (see figure 5):

Figure 5 cleary shows that an amplitude valley at the transition phase can only be observed within F3 (frame 14-18), whereas the amplitudes of F1 and F2 exhibit a continuous drop.

In the examples discussed above, a subsequent clime of the amplitude of all formants has been observed after the transition phase. As a next step towards assimilation, this subsequent clime of the amplitude is omitted. However, several modes of realization at this stage of assimilation can be observed.

3.4.Different sorts of amplitude modulation

At this stage of assimilation, the boundary between the two vowels is already difficult to define. The only clue to a possible boundary is provided by a change in the amplitude of one of the formants. In the following example (figure 6), a valley can be observed in the amplitude of F1 and F2 at frame 16, with a subsequent rise of both formant frequency amplitudes.

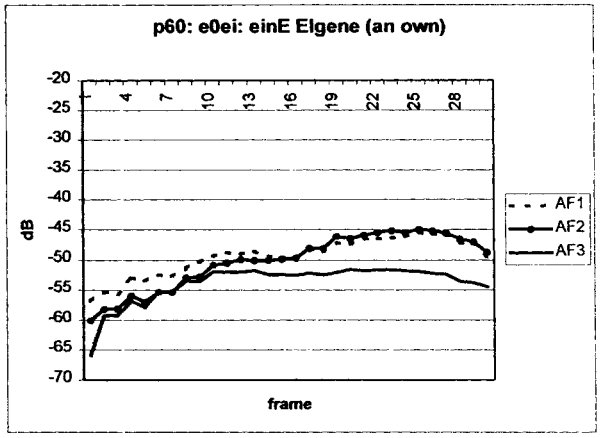

figure 6: frame by frame amplitude contour of F1, F2 and F3.

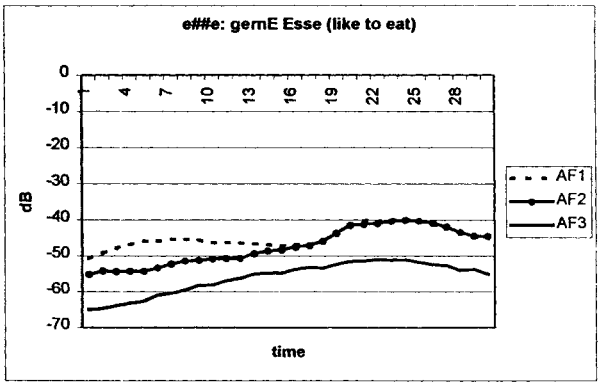

figure7: frame by frame amplitude contour of F1, F2 and F3.

The next example (figure 7) shows a sudden rise of both first and second formant frequency at frame 17, whereas the amplitude of F3 is raising continuously. In this example, segment boundary might be positioned at frame 17 at the sudden rise of the amplitudes of F1 and F2.

The whole sequence of the next example (figure 8) is produced in creaky voice, therefore, creaky voice can no longer function as a parameter for segment boundary determination. The only cue for a possible segment

boundary is provided by a sudden drop in amplitude of F1 at frame 14.

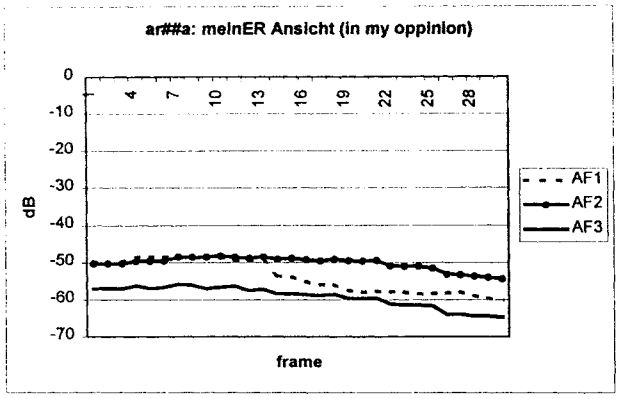

figure 8: frame by frame amplitude contour of F1, F2 and F3.

The modulation of the amplitude of only one formant is the last step before the two adjacent vowels assimilate totally. For the time being, no hierarchy with respect to the order of formant frequency amplitude assimilation could be determined.

3.5.Total assimilation.

Total assimilation with respect to all parameters discussed in the present study marks the endpoint of this gradual process. In this case, no cue to a possible boundary between the two adjacent vowels can be found any more (see figure 9):

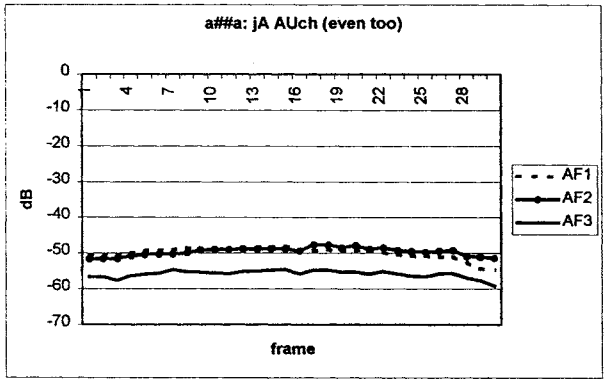

figure 9: frame by frame amplitude contour of F1, F2 and F3.

No movement can be observed neither with respect to formant frequencies or formant frequency amplitudes nor with respect to voice quality including fundamental frequency changes.

4.CONCLUSION

From the previous analysis, the following gradual steps of assimilation could be extracted:

1. The phase, in which the glottis is kept closed (glottal stop) is substituted by creaky voice. The segment boundary is therefore not as clear any more and it yields to a more or less blurred transition phase.
2. The irregularity in the periodicity of the waveform (creaky) is replaced more and more by modal phonation. Nevertheless, the transition phase might still be marked by a drop in fundamental frequency.
3. The whole sequence is realized in modal phonation, but the transition phase is made apparent by means of a valley of all three formant frequency amplitudes.
4. Where the amplitude valley is blurred, the segment boundary might be indicated by a sudden rise or drop of the amplitude of one of the formants.
5. The assimilation of formant might take place from stage 1 to stage 3.
6. The assimilation is completed, no indication of a segment boundary can be found any more.

6.REFERENCES

[1] L. Madelska and W. U. Dressler, *Postlexical Stress Processes and Their Segmental Consequences Illustrated in Polish and Czech.* In: B. Hurch and R. Rhodes (eds.) "Natural Phonology: The State of the Arts", de Gruyter, Berlin, *1996,* pp. 189-200.

[2] S. Moosmüller, "Hochsprache und Dialekt in Österreich", Böhlau, Wien, 1991.

[3] P. Helgason and K. J. Kohler, "Vowel Deletion in the Kiel Corpus of Spontaneous speech", *AIPUK* 30, pp. 115.157, 1996.

[4] D. R. van Bergem,"*Experimental Evidence for a Comprehensive Theory of Vowel Reduction",* Proc. EUROSPEECH'95, pp: 1319-1322, Berlin, 1995.

[5] R.J.J.H. van Son and L.C.W. Pols, "A Comparison between the Acoustics of Vowel and Consonant Reduction", *IFA Proceedings,* Vol. 20, pp. 13-25, 1996.

[6] K. Vollmer, "Koartikulation und glottale Transparenz", phonetik AIMS, Universität Stuttgart, 1997.

[7] D. Duez, "On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced stops", *Journal of Phonetics,* Vol. 23, pp. 407-427, 1995.

[8] S. Moosmüller, "*Diphthongs and the process of monophthongization in Austrian German: a first approach*", Proc. EUROSPEECH'97, pp. 787-790, Rhodos, 1997.

[9] W.U. Dressler, "Explainig natural phonology", *Phonology Yearbook,* Vol. 1, pp. 29-50, 1984.

[10] W.U. Dressler and S. Moosmüller, "Phonetics and Phonology: A sociopsycholinguistic framework", *Phonetica,* Vol. 48, pp. 135-148, 1991.

[11] W.A. Deutsch. and A. Noll, "S_Tools. Datenerfassung, Speicherung und Digitale Signalverarbeitung für Akustik, Lärm, Sprache und Musik", Austrian Academy of Sciences, Acoustics Research Department, Vienna, 1994.