Formant Trajectories in CV Clusters for German Stops

Tomasz Kuczmarski

INSTYTUT JĘZYKOZNAWSTWA, UNIWERSYTET IM. ADAMA MICKIEWICZA
AL. NIEPODLEGŁOŚCI 4, 61-874 POZNAŃ

`faqster@gmail.com`

Abstract

This paper presents an attempt to demonstrate the nonlinear progress of vowel formants in German CV clusters with stop sounds and identify it as a significant acoustic property characterizing the place of articulation of the stop consonant.

0 Introduction

Stops are consonantal sounds consisting of

> "a sequence of several articulatory and acoustic events. There is an initial phase in which a closure is made at a particular point along the vocal tract by manipulating the appropriate articulatory structure. This phase is followed by a time interval in which the vocal tract remains closed" (Stevens 1998)

and is characterised by no or very little energy in all frequency ranges in signals spectra. Eventually, the closure is released producing an interval of aperiodic noise.

Because of the above nature of stop consonants, three main cues for place of articulation might be identified. However, in the present paper only formant trajectories are considered. Nevertheless, the remaining two, being the spectral characteristics of the burst itself and some visual cues to the position of the tongue, lips, etc. (Jackson 2001, Stevens 1998), should not be underrated because they might provide a good source for comparison for future studies.

The choice of formant trajectories is justified by results of many past experimental studies by Grimm (1966), Öhman (1966) and Furui (1986), amongst others, where listeners were presented with backward-gated versions of CV syllables, originally composed of a stop consonant and a vowel. Results showed that formant transitions in remaining segments caused by coarticulation allowed listeners to classify the place of articulation of the gated consonant with a small error rate (Smits 2009).

Additionally, other studies where the acoustic attributes of presented consonant-vowel syllables were systematically manipulated, show that "stimuli with bursts only and no [formant] transitions were not consistently identified." (Stevens 1978)

In this study, I analyze six German stop consonants (/p/, /t/, /k/, /b/, /d/ and /g/) in three vowel contexts (/a/, /I/ and /U/) and expect to find some distinctive tendencies in vowel formant transitions as those schematically presented in (Fig. 1).
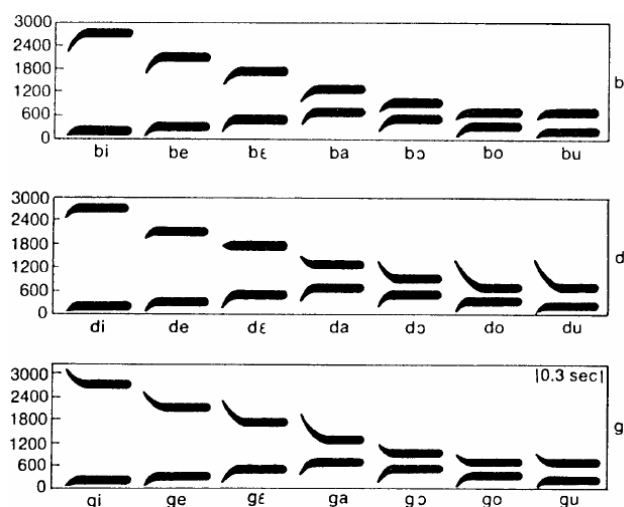
Figure 1: Schematic formant transitions in CV clusters.

1 Methods

An SWRK speech corpora, a part of SmartWeb project, was used for the purpose of the study. The corpus consists of recordings of a German female professional speaker and was designed for the purpose of speech synthesis. It contains a total of 92617 segments, among which 24999 are vowels, including 2358 occurrences of /a/, 2640 occurrences of /I/ and 1430 occurrences of /U/.

Given festival utterance structure for each sentence, features such as name and position in syllable of current and following segment and formant frequencies at nine equal time intervals for first three formants of the following segment (pos_in_syl, n.pos_in_syl, name, n.name, n.lisp_{f1, f2, f3}_{10, 20, 30, 40, 50, 60, 70, 80, 90}) were extracted for all phones that appeared in the corpora.

Table 1: Number of extracted CV clusters

	/a/	/I/	/U/
/p/	105	26	43
/t/	195	145	66
/k/	107	38	86
/b/	111	86	78
/d/	298	105	80
/g/	112	48	18

A simple perl script was used to sort out all CV clusters for C=/p,t,k,b,d,g/ and V=/a,I,U/ that occurred within one syllable. In result 1747 different CV clusters were extracted. Detailed numbers are presented in the table above (Tab. 1).

For each consonant-vowel cluster mean values of formant frequencies at nine equal time intervals of all occurrences were calculated. Data with normalized time was plotted on three charts in accordance to the vowel context. After a brief analysis of the preliminary results it turned out that in most of the cases there is a very high rate of formant frequency variation demonstrated by unnatural formant trajectories (Fig. 2, 3, 4).
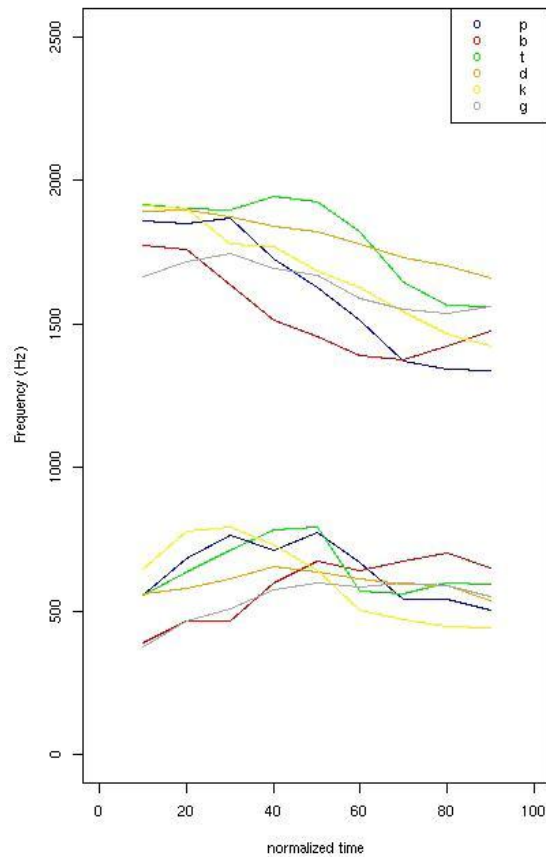
Figure 2: Preliminary results of formant trajectories for six German stops: /p/, /t/, /k/, /b/, /d/ and /g/ in /a/ clusters.
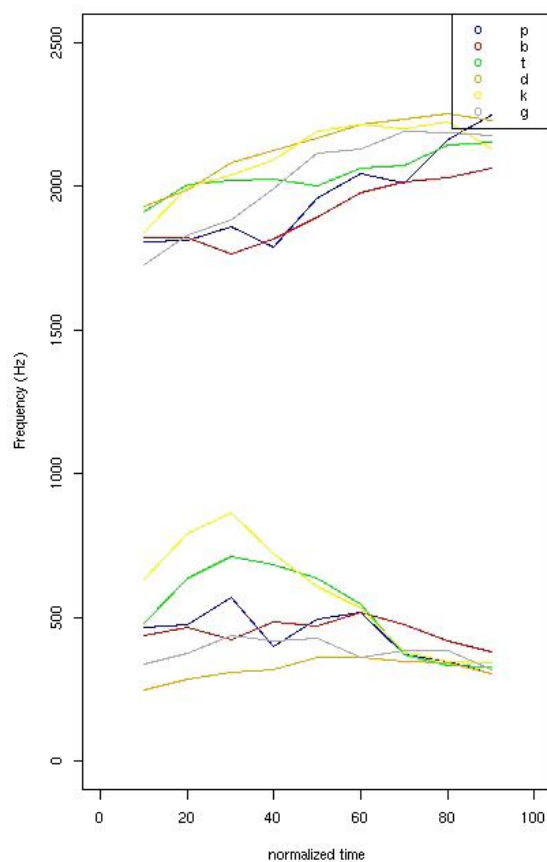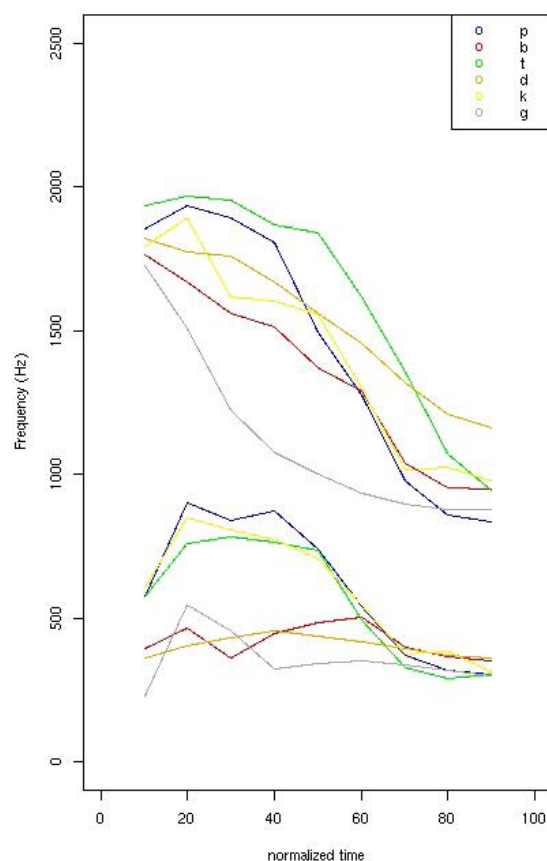

Figure 3: Preliminary results of formant trajectories for six German stops: /p/, /t/, /k/, /b/, /d/ and /g/ in /I/ clusters.

A simple assumption was made that formants in the middle part of the vowel have usual frequencies as in (Tab. 2), regardless of any transitions at the beginning.

On this basis a biased hand check of the data was performed and all records with unusual formant frequencies in the middle part of the vowel were deleted. The size of the resulting data set is presented in (Tab. 3).

Upto 20 (were possible) occurances of each cluster with smallest formant frequency confusion rate were hand-picked, except for the voiceless stop context clusters and /U/ clusters, where the whole set was ignored having unnatural values in almost every record. Mean values of formant frequencies were calculated again and plotted on charts with a normalized time scale. (Fig. 5, 6)

Figure 4: Preliminary results of formant trajectories for six German stops: /p/, /t/, /k/, /b/, /d/ and /g/ in /U/ clusters.

Table 2: Mean frequency of the first two formants for German /a/, /I/ and /U/ based on (Neppert & Petursson 1992)

	F1	F2
/a/	646	1301
/I/	363	2088
/U/	369	750

Table 3: Data set size after the hand check.

	F1	F2
/ba/	20	12
/da/	20	18
/ga/	20	19
/bI/	20	20
/dI/	20	20
/gI/	20	5

2 Results

In the raw, automatically extracted data set, especially at the beginning of vowels, some formant frequency values are unacceptably far from what could be expected in accordance to (Fig. 3). Some F1 values of /U/, for example, reach as high as what usually is /U/'s normal second formant frequency range. Yet, formants tend to stabilize at the end of segments leading to huge trajectories.
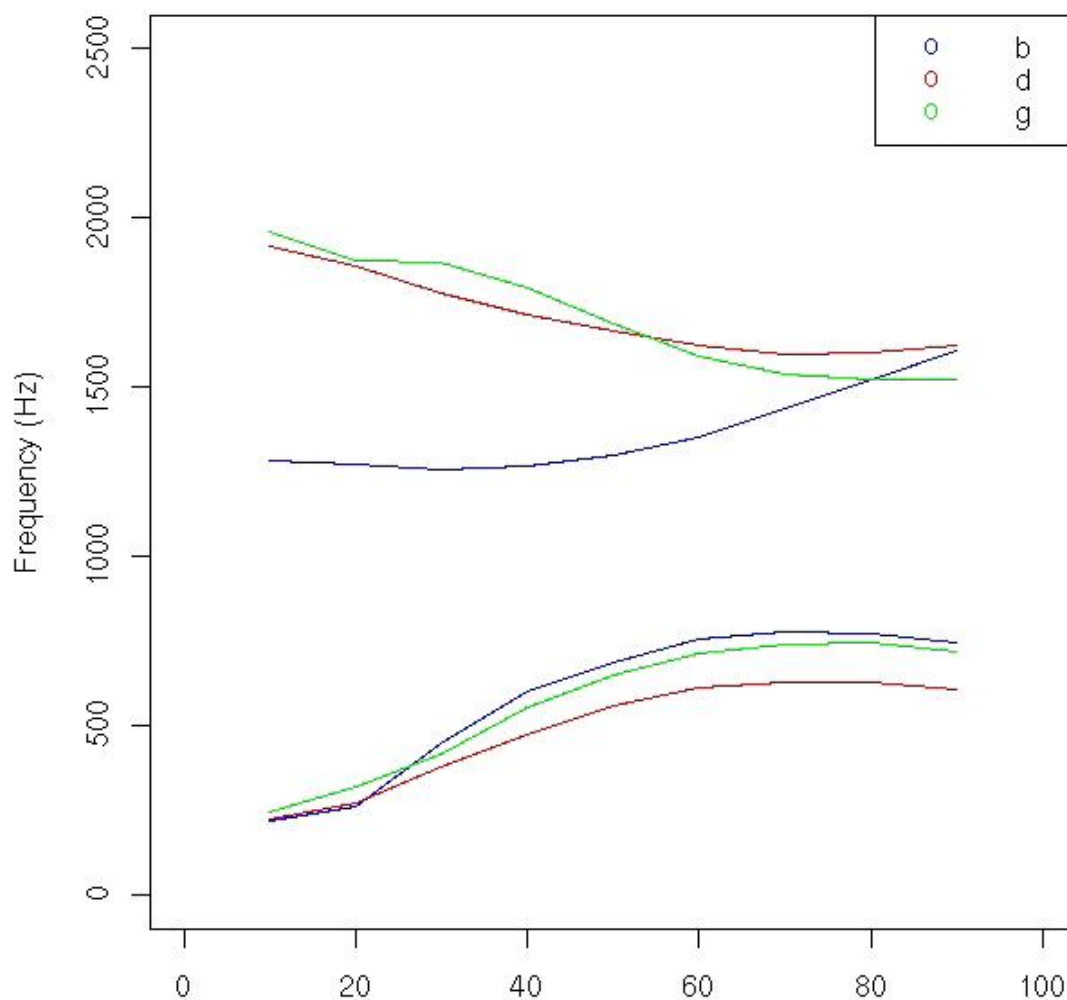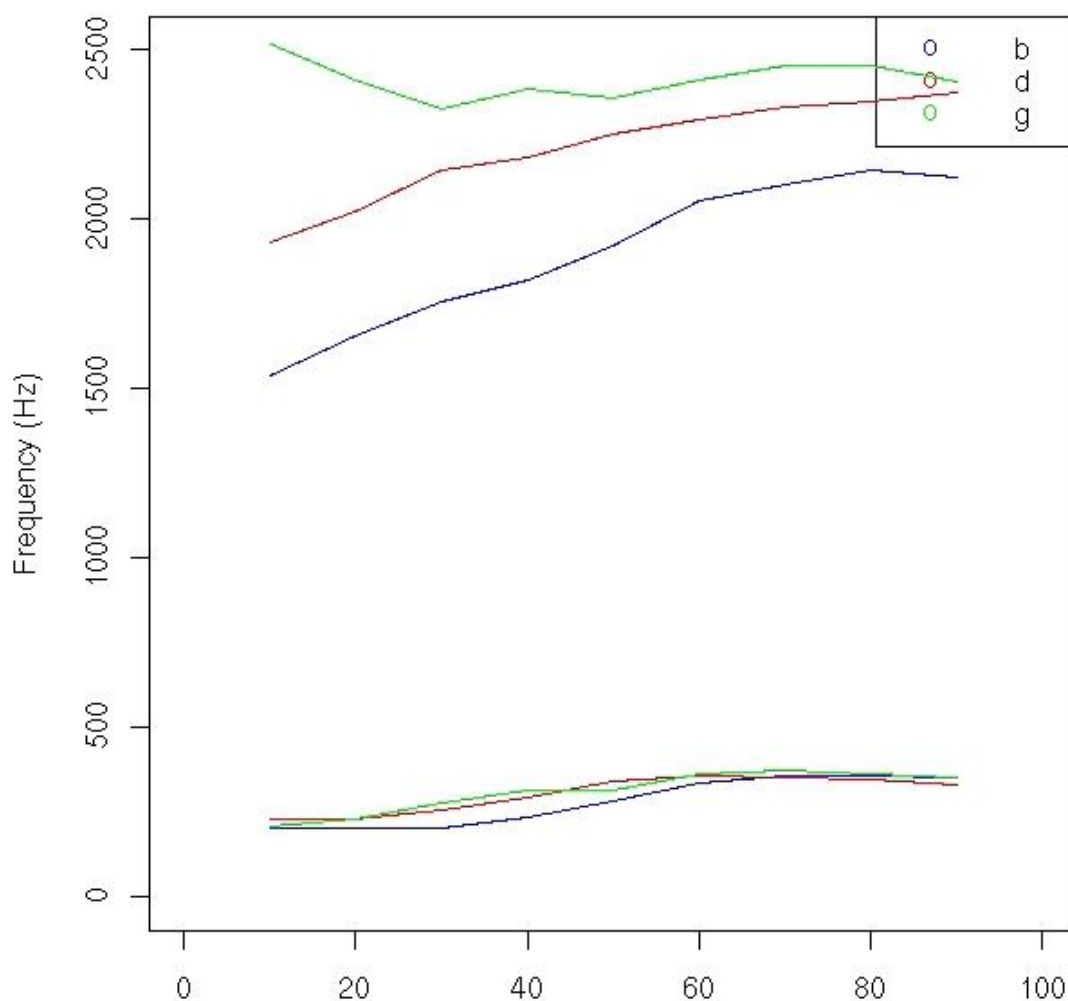

Figure 5: /a/ formant trajectories in /b/, /d/ and /g/ CV clusters in German.

Figure 6: /I/ formant trajectories in /b/, /d/ and /g/ CV clusters in German.

Hand adjustment eliminated improper values what resulted in smoothened mean of formant frequencies. Now they all have a rather stable run (Fig. 5, 6). Transitions are smaller but vivid. Both formants tend to drop down towards the voiced labial stop /b/ independently of the vowel. First formants, both of /a/ and /I/ similarly start rather low after the voiced alveolar stop /d/ and then increase to normal frequency ranges of the adequate vowels. However, the second formant of /I/, which normally occurs in higher frequencies (above 2000Hz) in /d/ context starts a bit lower and only than proceeds to its normal value. The second formant of /a/, on the other hand, which usually appears at around 1300Hz, after /d/ starts a bit higher. First formant frequencies of /a/ and /I/ are also both decreasing towards the voiced velar stop /g/. Both vowels second formants, even the one of /I/, start even higher in /g/ contexts to quickly plummet down to their natural frequency ranges.

3 Discussion

Preliminary results based on the raw data set were highly erratic because of false formant frequency values in the SWRK corpora. Those errors were probably caused by the algorithm used to extract formant frequencies from the signal itself and seem to have mistaken neighbouring formant values for each other what is clearly visible on the charts in (Fig. 2, 3, 4). For all three voiceless stop contexts first formant of /U/ starts where the second formant should be (Tab. 2) and falls down to its regular frequency only in the second half of the vowel. Second formant was correspondingly mistaken for the third formant. Formants were confused less often at the end of vowel segments and that generally caused the huge trajectories in most of the cases.

The error rate is highest for all voiceless stops as well as /U/ in general. Problems with the latter occurred probably because its F1 and F2 are both low and close to each other and even generally robust algorithms might have problems distinguishing them. Voiceless stops, in turn, might be problematic because of the positive VOT (Voice Onset Time) and *blurry* and indistinguishable formants at the beginning of the segment. Therefore, all voiceless stop clusters and /U/ clusters were ignored in this study.

A brief hand adjustment of the remaining data decreased the error rate but was evidently biased, being strongly dependent on subjective expectations. However, it helped demonstrate the general tendencies of vowel formant transitions in CV clusters with voiced stops in German.

Possibly, another similar study should be conducted in the future. Formant frequencies should be extracted using some dedicated technique that provides robust algorithms to deal with vowels in stop consonant contexts, as the one proposed by (Zheng 2004). That would help reduce data loss due to high error rates, which this study suffered from.

Bibliography

Jackson, P.J.B. 2001. Acoustic cues of voiced and voiceless plosives for determining place of articulation. in: *Proc. CRAC Workshop.* Aalborg, Denmark. pp. 19–22.

Neppert J., Pétursson, M. 1992. *Elemente einer akustischen Phonetik.* Hamburg: Helmut Buske Verlag.

Smits, R. 2009. *Human Consonant Recognition For Initial And Final Segments of VCV Utterances.* @: http://pitch.phon.ucl.ac.uk/home/shl10/roel/real.htm. Department of Phonetics and Linguistics, University College London.

Stevens, K. N. 1998. *Acoustic phonetics.* Cambridge: MIT Press.

Stevens, K. N., Blumstein S. E. 1978. Invariant cues for place of articulation in stop consonants. in: *J. Acoust. Soc. Am.*, vol. 64, no. 5, pp. 1358–1368.

Zheng, Y., Hasegawa-Johnson, M., Borys, S. 2004. Stop consonant classification by dynamic formant trajectory. in: *Interspeech-2004*, pp. 2481-2484.