Locus Equations Determination Using the SpeechDat(II)

Bojan Petek

Interactive Systems Laboratory
University of Ljubljana, Slovenia

Bojan.Petek@Uni-Lj.si

10.21437/Eurospeech.2003-207

Abstract

This paper presents a corpus-based approach to determination of locus equations for Slovenian language. The SpeechDat(II) spoken language database is analyzed first for all available target VCV contexts in order to yield candidate subsets for the acoustic-phonetic measurements. Only the VCVs embedded within judiciously chosen carrier utterances are then selected for the (F2 $_{vowel}$, F2 $_{onset}$) measurements. The paper discusses challenges, methodology, and results obtained on the 1000-speaker Slovenian SpeechDat(II) database in the framework of /VbV/, /VdV/, and /VgV/–based determination of locus equations.

1. Introduction

Locus equations are linear regression fits estimated on the second formant (F2) scatterplots that are formed by plotting two F2 measurements in CV transitions: (i) the onset frequency of the F2 transition immediately following the consonant release burst (F2 $_{onset}$, in Hertz, plotted along the ordinate) and (ii) the F2 at the vowel target position (F2 $_{vowel}$, in Hertz, plotted along the abscissa). Therefore, each <x,y> measurement represents a simplified parametrization of the F2 transition. When a scatterplot of measurements is formed for a given stop consonant across various vowel contexts, it was shown that the <x,y> coordinates for a given stop place category (the allophones) exhibit a highly linear distribution across this F2-defined acoustic space. Research reported by Chennoukh [2], Fowler [3], Lindblom [4], and Sussman et al. [5-9] has thoroughly explored the acoustic properties of F2 in CV transitions and addressed various aspects of the non-invariance problem of the stop place representation and perception.

This paper presents a corpus-based approach to locus equations determination using the Slovenian SpeechDat(II). Main motivation behind this goal was (i) to determine the locus equations for Slovenian language and (ii) to explore the challenges in locus equation determination on a speech technology database, i.e., by using data quality that might not be best suited for the acoustic phonetic research.

2. Methodology

The Slovenian SpeechDat(II) corpus [11] was analyzed first for all target VCVs where C denotes the Slovenian /b,d,g/ stops and V the Slovenian vowels /i,e,E,a,O,o,u/ [10]. Each nonempty SpeechDat(II) VCV subset was analyzed in order to select particular VCVs embedded in candidate carrier utterances. In this paper up to four speakers (two female and two male) have been selected per specific VCV and were, whenever possible, sampled to contain lexically identical carrier utterances. This yielded a pool of Slovenian /VbV/, /VdV/, and /VgV/ subsets that were analyzed for the F2 $_{onset}$ and F2 $_{target}$ frequencies using *Praat* tool, version 4.0.38 [1]. All measurements judged to be consistent were entered into the final scatterplot for linear regression analysis for each of the /VbV/, /VdV/, and /VgV/ subsets that served for the locus equations determination.

3. Slovenian SpeechDat(II) Corpus

Slovenian SpeechDat(II) was recorded over the fixed telephone network using the ISDN interface. It contains recordings of 1000 speakers and represents to date one of the largest spoken language resources available for the development of Slovenian spoken language applications [11].

The corpus contains recordings of the application words (e.g., *meni* ‘menu’, *prekliči* ‘cancel’, *ustavi* ‘stop’, etc.), sequence of isolated and connected digits (e.g., PINs, telephone and credit card numbers), dates, spelled words and phrases, directory assistance names (e.g., city names, company/institution names), questions, and phonetically rich words and sentences. The corpus also includes a limited amount of spontaneous speech.

Recordings were predominantly taken in home and office environments (73% and 21%, respectively), while 6% of the calls originated from other acoustic environments. Unlike SpeechDat(II) databases collected for other languages, the Slovenian database contains more noise. Procedure reported in the database validation phase for the signal to noise ratio (SNR) estimation involved computation of mean square energy of 10 ms windows. 5% of the windows with the lowest energy were judged to contain line noise. The SNR for each file was obtained by dividing the mean energy over all windows by the mean energy of the 5% sample mentioned above. Table 1 summarizes the SNR results reported for the Slovenian SpeechDat(II). As shown the most of estimated SNRs are within the range of 20 – 35 dB.

Table 1: The SNR distribution reported for the Slovenian SpeechDat(II) [11].

SNR [dB]	# of occurences
0 – 5	132
5 – 10	239
10 – 15	770
15 – 20	2914
20 – 25	11987
25 – 30	20127
30 – 35	6704
35 – 40	116

4. Results

4.1. Slovenian SpeechDat(II) VCV corpus analysis

The first research step involved a distribution analysis of the /VbV/, /VdV/ and /VgV/ contexts that were used to determine a list of candidate carrier utterances for acoustic phonetic measurements. The SpeechDat(II) contains only orthographic lexical transcriptions. Since the database is typically used in the development of speech technology applications, details on speech and non-speech audible acoustic events are also transcribed. Therefore, the VCV frequency distribution results reported in the following are derived on the basis of orthographic lexical transcription.

Figure 1 shows that the most frequent VCV (for the locus equations determination) in the 1000-speaker Slovenian SpeechDat(II) is /ede/ (with 5157 occurences), followed by /ide/ (1756 occurences), /eda/ (454), /oda/ (439), /ada/ (427), /ega/ (408), /udi/ (399), /aga/ (319), and /edi/ (316). The most frequent /VbV/ was found to be /obe/ (277 occurences), followed by /obo/ (252), /ibo/ (149), /obi/ (143) and /abi/ (129 occurences). For /VgV/ the most frequent context in the Slovenian SpeechDat(II) is /ega/ (408), followed by /aga/ (319), /ogo/ (164), /ugi/ (161) and /oga/ (155 occurences).

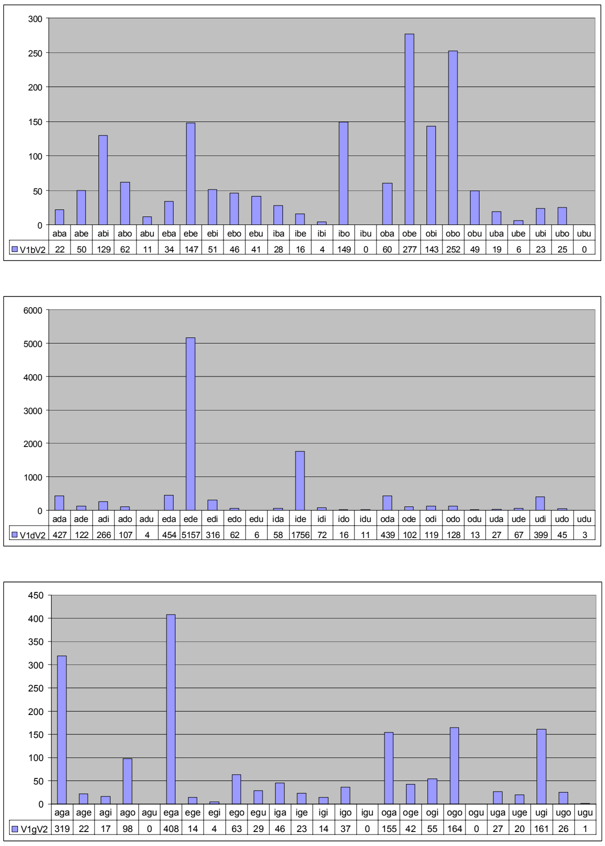

Figure 1: Number of occurrences of all /VbV/, /VdV/, and /VgV/ contexts in the Slovenian SpeechDat(II).

From each of the VCV SpeechDat(II) subsets, a sample of 4 speakers (2 female and 2 male) was further selected that typically contained target VCVs embedded in the lexically identical carrier utterances. Table 2 illustrates the results of this procedure for the case of /oda/. All 4 utterances were judiciously selected from the SpeechDat(II) sub-corpus which contained the total of 439 utterances previously extracted from the whole 1000 speaker corpus (see Figure 1).

Table 2: Details on the selected SpeechDat(II) carrier utterances for the /oda/ VCV. [sta]and [spk] denote the stationary and speaker noise, respectively.

Gender	Dialect region	/oda/ carrier utterances
F	LJU	informacije iz elektronske baze podatkov so brezplačne
F	ROV	informacije iz elektronske baze podatkov so brezplačne
M	STA	informacije iz elektronske baze podatkov so brezplačne
M	DOL	[sta] [spk] informacije iz elektronske baze podatkov so brezplačne [spk]

The /oda/ realizations in this case are positioned within the selected carrier utterances *informacije iz elektronske baze podatkov so brezplačne,* 'information from the electronic database is free of charge'.

Table 2 also reveals that the selected 2 female and 2 male speakers originate from different dialect regions. Therefore, the locus equations results reported in this paper represent a pooled version across the various dialect regions of Slovenia.

This heuristics was also chosen for the /VbV/- and /VgV/-based locus equations determination.

4.2. F2 $_{onset}$, F2 $_{vowel}$ measurements

Acoustic phonetic measurements were performed within the manually positioned VCV boundaries in the SpeechDat(II) carrier utterances. Figure 2 illustrates a typical /VbV/ example of the F2 $_{onset}$, F2 $_{vowel}$ frequency measurements at the CV transition in /iba/ using the *Praat* tool.

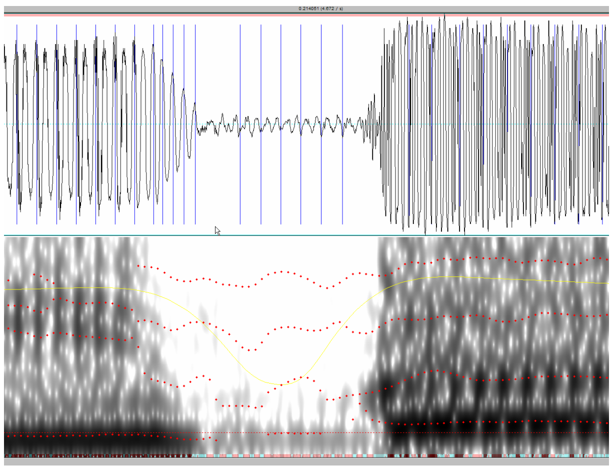

Figure 2: Example SpeechDat(II) CV analysis of the F2 $_{onset}$, F2 $_{vowel}$ in /iba/ using Praat.

Dotted lines in the spectrogram represent the formant trajectories estimated by *Praat*. Figure 3 illustrates the F2 $_{onset}$, F2 $_{vowel}$ frequency measurement at the CV transition in /ada/. This segment was embedded in the carrier sentence *naša vlada priporoča čim poznejše sklepanje zakonskih zvez* ‘our government recommends to get joined in marriage as late as possible’.

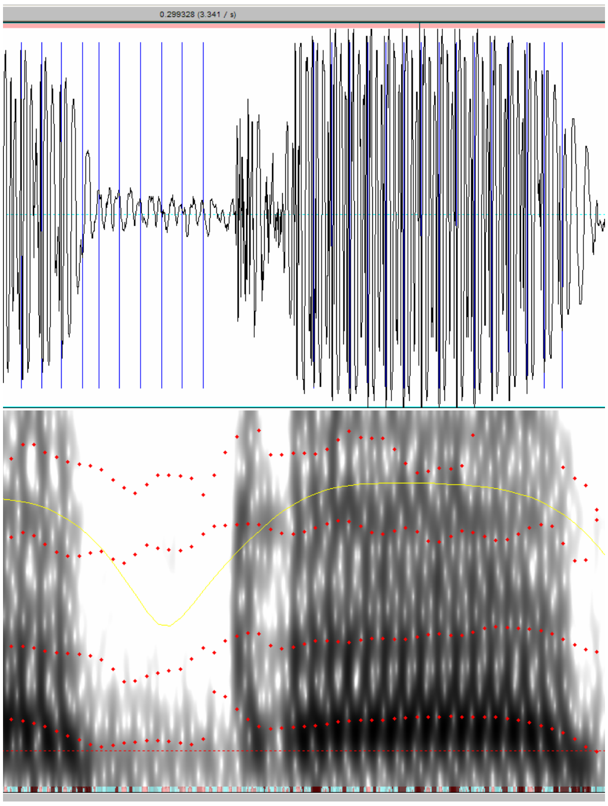

Figure 3 Example SpeechDat(II) CV analysis of the $F2_{onset}$, $F2_{vowel}$ in /ada/ using Praat.

The solid line in Figure 3 represents the energy contour and was used as a guide in the $F2_{vowel}$ determination. Figure 4 shows a typical example of the $F2_{onset}$, $F2_{vowel}$ frequency measurement for the case of /oga/.

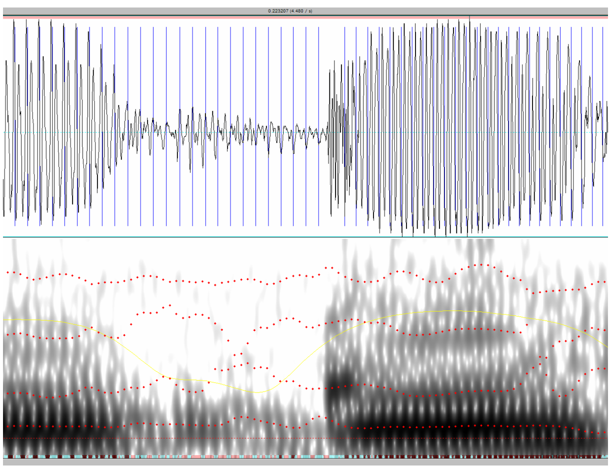

Figure 4: Example SpeechDat(II) CV analysis of the $F2_{onset}$, $F2_{vowel}$ in /oga/ using Praat.

The carrier sentence in this case was *Sonja, koga se bojiš* ‘Sonia, who are you afraid of ’.

4.3. Locus Equations

Locus equations results are shown in Figure 5.

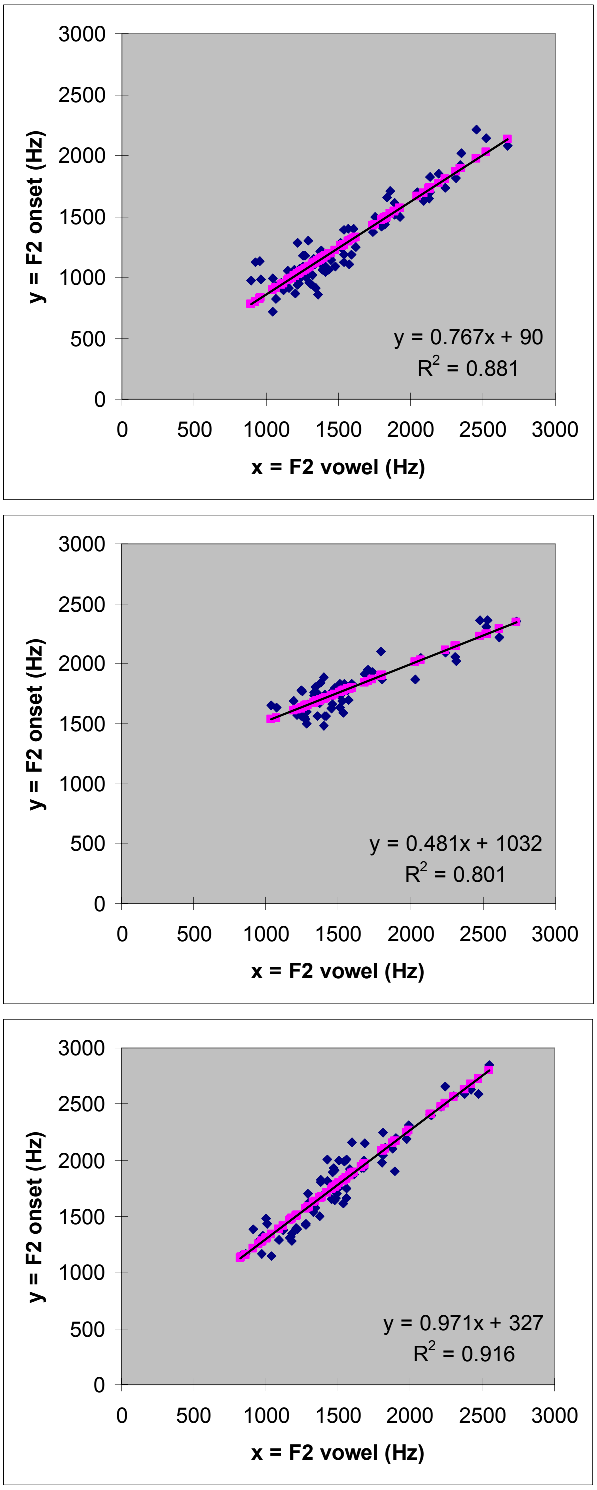

Figure 5: Locus equations results obtained on the Slovenian SpeechDat(II) for the selected /VbV/, /VdV/, /VgV/ contexts, respectively.

5. Discussion

Figure 5 shows that the regression statistics for locus equations determination on the Slovenian SpeechDat(II) yielded the R^2 values of 0.88, 0.80, and 0.92, for /b,d,g/, respectively. Standard errors (SEs) of estimate were 115 Hz for /b/, 99 Hz for /d/, and 123 Hz for /g/. Relatively low R^2 values (high SEs) can be explained by the fact that the SpeechDat(II) database is not ideally suited for the acoustic phonetic measurements since its main purpose is to support the development of spoken language applications in telephony. All $F2_{onset}$, $F2_{vowel}$ measurements were checked for consistency, yet there were many instances where the F2 formant estimation was problematic due to the unclear formant structure and/or low quality of the transmission channel. As mentioned earlier, clear F2 measurement outliers were removed from the final linear regression analysis.

The locus equations determined on the Slovenian SpeechDat(II) and expressed in Hertz for /b,d,g/ are $y = 0.767x + 90$, $y = 0.481x + 1032$, and $y = 0.971x + 327$, respectively.

Relatively high appears the value of slope coefficient for the Slovenian /g/. In English, for example, the labials have been found to have the steepest regression function, followed by velars, then alveolars. For Slovenian, the determined locus equation for /g/ appears to be quite different from the English one. Additionally, two allophonic groupings observed in English, i.e., $[g]_{velar}$ for /g/ preceding the back vowels and $[g]_{palatal}$ for /g/ preceding the front vowels were not observed in the case of Slovenian.

6. Conclusions

This paper explored an alternative possibility in determination of locus equations, i.e., by using the 1000-speaker Slovenian SpeechDat(II) spoken language corpus which is primarily targeted to support the speech technology applications in telephony. Since orthographic transcription in Slovenian language follows the phonetic one quite closely, the corpus was analyzed and judiciously sub-sampled to yield all VCVs suitable for the task of locus equations determination.

Research presented on locus equations for the Slovenian language and the reported results (Figure 5) are important for a cross-linguistic analysis. The methodology presented used sample sets of considerable size (Figure 1) and yielded novel insights despite the F2 measurement challenge related to the SpeechDat(II) ISDN data quality. Complementary work could concentrate on including a larger selection of speakers for the sub-sampled VCVs. Another interesting research avenue could be to select the speakers of standard Slovenian to build a high quality database for acoustic-phonetic research and to validate the Slovenian locus equation metric cross-linguistically.

7. Acknowledgements

Research reported in this paper was supported in part by the Slovenian-Norwegian bilateral project "Speech Analysis Techniques in Phonetics" granted by the Ministry of Education, Science and Sport of the Republic of Slovenia, the EU TIST COST Action 277 "Non-Linear Speech Processing", and made viable by the Socrates/Erasmus multilateral agreement D-IV-1/99-JM/Kc.

8. References

[1] Boersma, P., and Weenink, D., "Praat: doing phonetics by computer", http://www.fon.hum.uva.nl/praat/ (March 2003).

[2] Chennoukh, S., Carre, R., and Lindblom, B., "Locus equations in the light of articulatory modeling", J. *Journal of the Acoustical Society of America,* 102(4): 2380-2389, 1997.

[3] Fowler, C. A., "Invariants, specifier, cues: An investigation of locus equations as information for place of articulation", *Perception and Psychophysics,* 55: 597-610, 1994.

[4] Lindblom, B. "On vowel reduction", The Royal Institute of Technology, Speech Transmission Laboratory, Rep. 29, Stockholm, 1963.

[5] Sussman, H. M., McCaffrey, H. A., and Matthews, S., "An investigation of locus equations as a source of relational invariance for stop place categorization", *Journal of the Acoustical Society of America,* 90(3): 1309-1325, 1991.

[6] Sussman, H.M., Hoemeke, K.A., and Ahmed, F. S., "A cross-linguistic investigation of locus equations as a phonetic descriptor of articulation", *Journal of the Acoustical Society of America,* 94: 1256-1268, 1993.

[7] Sussman, H.M., Fruchter, D., and Cable, A., "Locus equations derived from compensatory articulation", *Journal of the Acoustical Society of America*, 97: 3112-3124, 1995.

[8] Sussman, H.M., Fruchter, D., Hilbert, J., and Sirosh, J., "Linear correlates in the speech signal: The Orderly Output Constraint", *Behavioral & Brain Sciences*, 21: 241-299, 1998.

[9] Sussman, H.M., "Representation of phonological categories: a functional role for auditory columns", *Brain and Language*, 80: 1–13, 2002.

[10] Šuštaršič, R., Komar, S., and Petek, B., "Illustrations of the IPA: Slovene", *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*, Cambridge University Press, 135-139, 1999. (ISBN: 0521652367).

[11] http://www.elda.fr/cata/speech/S0056.html (March 2003).