Locus Equation Metric Applied to Slovenian

Bojan Petek
Interactive Systems Laboratory
University of Ljubljana, Slovenia
E-mail: Bojan.Petek@Uni-Lj.si

ABSTRACT

This paper presents results on the determination of locus equations for Slovenian language. The 1000-speaker Slovenian SpeechDat(II) spoken language database is searched first for all target VCV contexts in order to yield candidate subsets for the acoustic-phonetic measurements. Next, only VCVs embedded within the judiciously chosen carrier utterances are selected for the (F2 vowel, F2 onset) measurements in the framework of /VbV/, /VdV/, and /VgV/–based determination of locus equations. The paper concludes with the discussion of results from the cross-linguistic perspective.

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 stop place representation and perception. Sussman et al. [6] also compared the locus equations cross-linguistically. Their analysis involved Arabic, English, Swedish, Thai, and Urdu languages. This paper extends the cross-linguistic analysis of locus equations to include Slovenian. Main motivation behind this goal was (i) to determine the locus equations for Slovenian language and (ii) to compare the results obtained on the SpeechDat(II) speech technology database cross-linguistically.

The paper is organized as follows. Section 2 presents the methodology used in the locus equation determination for Slovenian. Section 3 presents major characteristics of the spoken language corpus from which utterances for the acoustic-phonetic measurements were taken. Section 4 reports results of the VCV corpus analysis from the perspective of the most frequent /VbV/, /VdV/, and /VgV/ contexts present in the Slovenian SpeechDat(II) database. This section also provides insights into the selection procedure of VCV carrier utterances from the Slovenian SpeechDat(II) and illustrates a representative challenging F2 measurement example in the locus equation determination. Finally, the Sections 5 and 6 report results on the locus equations for Slovenian and compare them cross-linguistically.

2. METHODOLOGY

The Slovenian SpeechDat(II) spoken language 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 the 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. CORPUS CHARACTERISTICS

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].

This 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.

ISBN 1-876346-48-5 © 2003 UAB

Recordings were predominantly made 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.

Figure 1 summarizes the SNR results reported for the Slovenian SpeechDat(II) [11]. As shown the predominant SNRs range of calls is between the 20 – 35 dB.

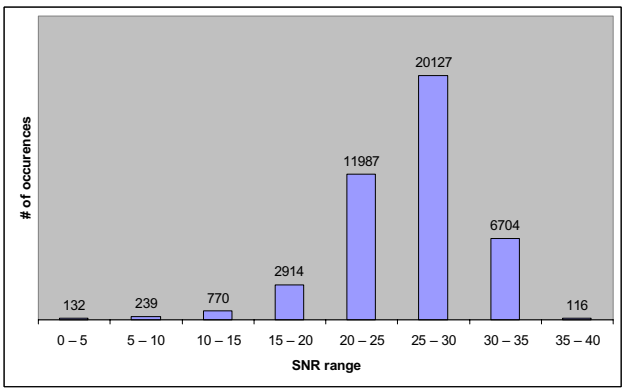

Figure 1: The SNR distribution reported for the Slovenian SpeechDat(II) [11].

4. RESULTS

The first step involved a distribution analysis of the /VbV/, /VdV/ and /VgV/ contexts that were used to determine a list of candidate carrier utterances for the acoustic phonetic measurements. The Slovenian SpeechDat(II) does not include manual phonetic transcriptions. Since the database is typically used in the development of speech technology applications, details on some speech and non-speech audible acoustic events are also transcribed (eg., [spk] and [int] denote the speaker and the intermittent noise, respectively) . Therefore, the VCV frequency distribution results reported here are derived on the basis of orthographic lexical transcriptions.

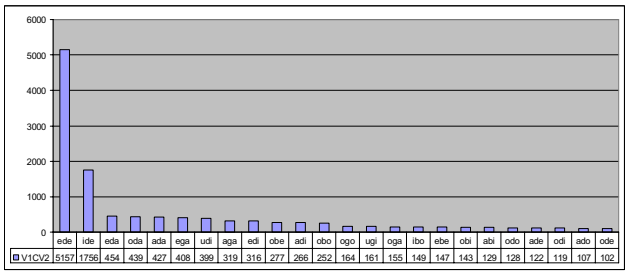

Figure 2*: The first 24 most frequent /VbV/, /VdV/, and /VgV/ contexts in the Slovenian SpeechDat(II).*

Figure 2 shows that the most frequent VCV (for the task of 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). The /e/ and /E/ (as well as /o/ and /O/) data in the VCV analyses were pooled.

4.1 CARRIER UTTERANCE SELECTION

From each of the VCV SpeechDat(II) subsets, a sample of 4 speakers (2 female and 2 male) was further selected that typically contained the target VCVs embedded in lexically identical carrier utterances.

Gender	Dialect region	/ibo/ carrier utterances
F	LJU	pozdravljeni, ali prispe brzovlak iz Maribora pred deseto uro
F	DOL	pozdravljeni, ali prispe brzovlak iz Maribora po pred deseto uro [spk]
M	STA	pozdravljeni, ali prispe brzovlak iz Maribora pred deseto uro [int]
M	DOL	pozdravljeni, ali prispe [spk] brzovlak iz Maribora pred deseto uro

Table 1: *Details on the selected SpeechDat(II) carrier utterances for the /ibo/ VCV.*

Table 1 illustrates the results of this procedure for the case of /ibo/. All 4 utterances were judiciously selected from the SpeechDat(II) sub-corpus which contained the total of 149 utterances previously extracted from the whole 1000 speaker corpus (see Figure 2). The /ibo/ realizations in this case are positioned within the selected carrier utterances *pozdravljeni, ali prispe brzovlak iz Maribora pred deseto uro*, 'hello, does the fast train arrive from Maribor before 10 o'clock'.

Another illustrative example of carrier utterance selection is given in Table 2. In this case the sub-corpus contained only 29 candidate utterances within the whole SpeechDat(II) corpus. The /egu/ VCVs are positioned within the carrier sentences *v tem snegu boš hodil najmanj sedem ur* 'in this snow you will walk for at least seven hours'.

Tables 1 and 2 also show that the selected female and male speakers originate from the various dialect regions. Therefore, the locus equations results reported in this paper represent a pooled version across the dialect regions of Slovenia.

Gender	Dialect region	/egu/ carrier utterances
F	LJU	v tem snegu boš hodil najmanj sedem ur [spk]
F	PRI	v tem snegu boš hodil najmanj sedem ur
M	DOL	v tem snegu boš hodil najmanj sedem ur
M	KOR	v tem snegu boš hodil najmanj sedem ur

Table 2: *Details on the selected SpeechDat(II) carrier utterances for the /egu/ VCV.*

This heuristics was also chosen for other /VbV/- , /VdV/-, and /VgV/-based carrier utterance extraction for the 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 3 illustrates a representative problematic VCV example of the F2 onset, F2 vowel frequency measurements at the CV transition in /ega/ using the Praat tool (ie., an example of a sloppy speaking style).

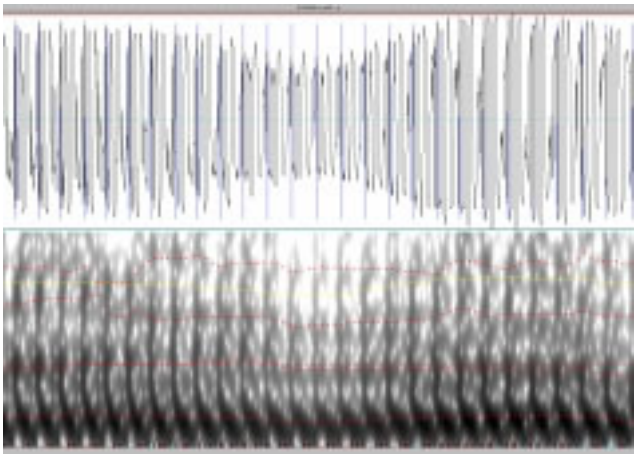

Figure 3: Representative problematic example of SpeechDat(II) CV analysis of the F2 $_{onset}$, F2 $_{vowel}$ in /ega/ using *Praat*.

Dotted lines in the spectrogram represent the formant trajectories estimated by *Praat*. This segment was embedded in the carrier sentence *je to količina denarja porabljenega za oglaševanje?* 'is this the amount of money spent for advertisements?'.

5. LOCUS EQUATIONS RESULTS

Locus equations results are shown in Figure 4. 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.

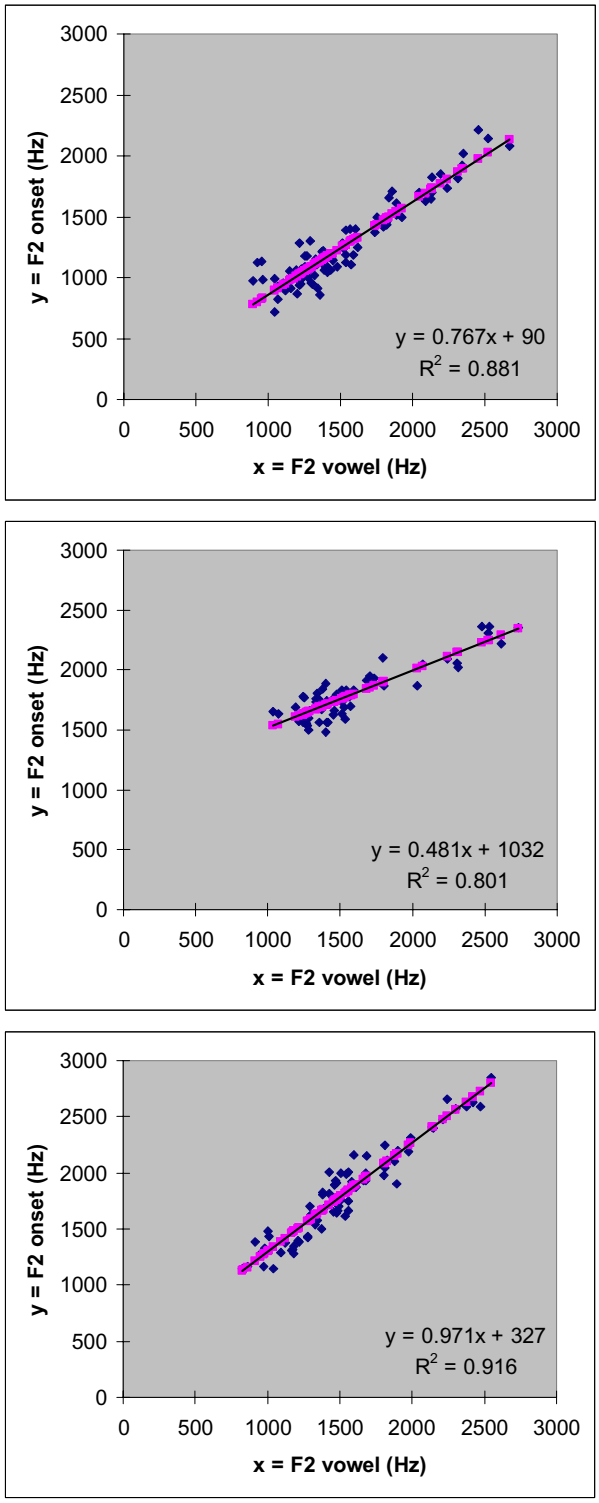

Figure 4: Locus equations results obtained on the Slovenian SpeechDat(II) for the selected /VbV/, /VdV/, /VgV/ contexts, respectively.

ISBN 1-876346-48-5 © 2003 UAB

6. DISCUSSION AND CONCLUSION

Figure 4 shows that the regression statistics for locus equations determination on the Slovenian SpeechDat(II) yielded the R2 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 R2 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.

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 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.

Table 3 positions the Slovenian locus equation results within the cross-linguistic comparison of [6]. Due to relatively low suitability of the Slovenian SpeechDat(II) database for the acoustic-phonetic measurements, however, it may be necessary to validate the results using the experimental setup that would involve the construction of a new standard Slovenian database. Such a large (preferably multimodal, high quality, annotated, validated and properly standardized) spoken language database could prove essential in the advanced fundamental research and in the development of human language technology. It could also be used to further develop the concept of portability of human language technologies [12] that builds upon the technological proximity and specificity of languages.

Language	/b/		/d/		/g/	
	k	y0	k	y0	k	y0
Arabic	0.77	206	0.25	1307	0.92	229
English	0.87	106	0.43	1073	0.66	807
Slovenian	0.77	90	0.48	1032	0.97	327
Swedish	0.63	487	0.32	1096	0.95	360
Thai	0.7	228	0.3	1425	-	-
Urdu	0.81	172	0.5	857	0.97	212

Table 3: *Locus equations coefficients for Slovenian and for different languages reported in [6].* k *and* y0 *denote the slope and the y intercept, respectively.*

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 was made viable by the Socrates/Erasmus multilateral agreement D-IV-1/99-JM/Kc.

REFERENCES

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

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

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

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

[5] H. M. Sussman, H. A. McCaffrey, and S. Matthews, “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] H. M. Sussman, K. A. Hoemeke, and F. S. Ahmed, “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] H. M. Sussman, D. Fruchter, and A. Cable, "Locus equations derived from compensatory articulation", *Journal of the Acoustical Society of America*, 97: 3112-3124, 1995.

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

[9] H. M. Sussman, “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”, in 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).

[12] http://www.lrec-conf.org/lrec2002/lrec/wksh/Portability.html (March 2003).