IDENTIFICATION OF REGIONAL VARIANTS IN THE STANDARD SLOVENIAN SPEECH

Bojan Petek
University of Ljubljana
Faculty of Natural Sciences and Engineering
Snezniska 5, 1000 Ljubljana, Slovenia
Bojan.Petek@uni-lj.si

ABSTRACT

Slovenian is a Slavic language of 2 million population. Its dialectal structure is very rich and exhibits 7 major dialect groups including more than 40 dialects. This paper addresses pronunciation variants of Slovenian digits that were collected over the fixed network all across Slovenia. Perceptual and acoustic phonetic experiments included measurements of the time normalized formant dynamics for the most confusable digits that were identified in series of isolated digit recognition experiments. Specifically, the main research avenue was to access the frequency and type of variation from the Standard Slovenian digit pronunciation. Recognition results yielded that the most confusable digits are dve `two´ /d v e/, pet `five´ /p e t/ and devet `nine´ /d ε v e t/. Perceptual and acoustic phonetic analyses showed that most speakers replied in the Standard Slovenian. Only two major dialectal classes were identified and characterized as diphthongizations of /e/ in the Standard Slovenian pronunciation.

1. INTRODUCTION

Voice interactive telephone services represent an important part in the forthcoming information age. In many interactive applications digits and digit strings play a key role. They are used to indicate date and time, credit card and account numbers, money amounts, and may be used for navigation through the interactive services. It is important to understand their pronunciation variations in a dialectally rich language such as Slovenian where more than 40 dialects in 7 dialect groups could be identified. Therefore, this paper addresses the Slovenian digit pronunciation patterns using a pool of speaker database collected all across Slovenia.

The paper is organized as follows. Section 2 outlines the adopted methodology and Section 3 reports the results. First, the speech database was perceptually analyzed for pronunciation variations and later examined for confusability of digits from the automatic speech recognition (ASR) point of view (Section 3.1). Second, using the insights from the ASR experiments, further perceptual and instrumental acoustic phonetic analyses addressed the variation of Slovenian digit pronunciation (Section 3.2). Measurements included analyses of the time normalized formant trajectory contours (Section 3.3). The purpose of these experiments was to investigate if perceptual evidence of the regional differences in pronunciation of digits could be modeled by the differences in formant dynamics. Such differences could potentially address refinement of the most confusable set of digit models thereby increasing the robustness of ASR.

2. METHODOLOGY

In the database acquisition phase a pool of 1000 speakers all across Slovenia was asked to pronounce ten Slovenian digits (0-9) and three control words (ja `yes´, ne `no´, stop `stop´) over the telephone line. 780 speakers responded positively to the initiative. Each word in a speaker record was embedded in 200ms and 1s leading and trailing telephone channel characteristic. This yielded a telephone signal database of 4 hours and 44 minutes sampled at 8kHz sampling rate.

All prompts for digit pronunciations were given in Standard Slovenian so most of informants adhered to reply in the same manner. Nevertheless, many speaker records exhibited a high degree of dialectal variation that was in primary focus of investigation reported here.

In order to study the pronunciation patterns of Slovenian digits a pool of 300 speaker records were selected for further analysis. Digits were first investigated for confusability from the ASR point of view. Gender specific 10-state continuous density word level HMM models were trained using up to 14 mixtures per state on a 200 speaker training set. Confusion matrix of the best test set recognition performance using another 100 speakers was then examined for the most confusable digit pairs.

Both training and test set digits belonging to the most confusable word sets were analyzed further perceptually and instrumentally. Perceptual and acoustic phonetic analyses identified pronunciation variants, i.e. each digit pronunciation was classified into the standard or dialectal class. Each dialectal digit pronunciation was further subdivided into one of the broad dialectal variants. Instrumental analysis included measurements of normalized F1 and F2 formant trajectory contours that were sampled at 8 equidistant points.

3. RESULTS

3.1. ASR experiments

Series of experiments using continuous density HMM word-level models yielded the best recognition accuracy of 97.3%. Confusion matrix analysis of the best test set performance identified that the most confusable digit pairs were tied to the digit /d ε v e t/ devet (9). These pairs were {/p e t/ pet (5), devet (9)}, {/d v e/ dve (2), devet (9)} and {tri /t r ɪ/ (3), devet (9)}. In the present study, Slovenian digits dve, pet, devet were selected for further perceptual and instrumental analyses.

3.2. Perceptual and instrumental analyses

Results of the perceptual and instrumental analyses for the digit dve `two´ are summarized in Table 1. This digit actually exhibited two consistent Standard Slovenian pronunciations (denoted by S1 /d v e/ and S2 /d v ɐ/). Given that the acoustical prompts to informants were given as "Reci /d v e/" `Say /d v e/´ one cannot conclude that the preferred Standard Slovenian pronunciation of the digit dve is /d v e/. Example Standard Slovenian pronunciation of the digit dve is given in Figure 1 [3,4,6].

Pronunciation type	Pronunciations
S1	79%
S2	2%
D1	9%
D2	6%
D3	4%

Table 1. Pronunciation variants of the Slovenian digit dve `two´.

Three broad dialectal pronunciation variants of digit dve could be identified in the database of 300 speakers. They are denoted as D1 to D3.

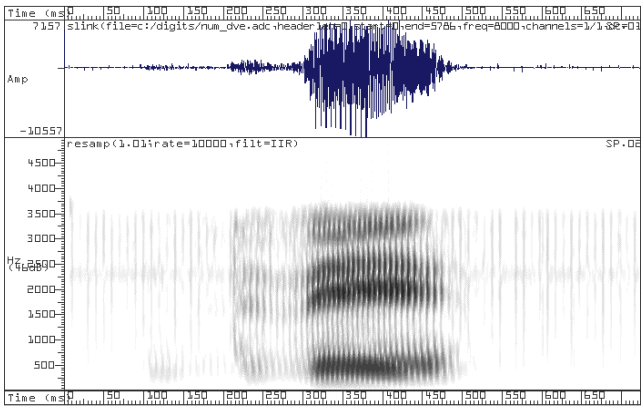

Figure 1. Example of the Standard Slovenian pronunciation of digit dve `two´ (S1) [SOUND 0340_01.wav].

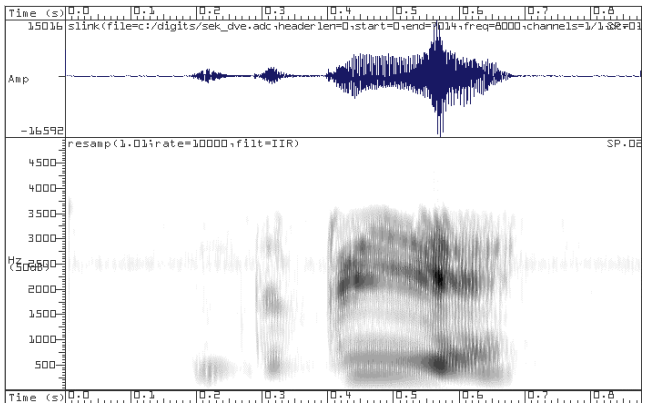

Figure 2. Example of the D1 dialectal pronunciation of Slovenian digit dve [SOUND 0340_02.wav].

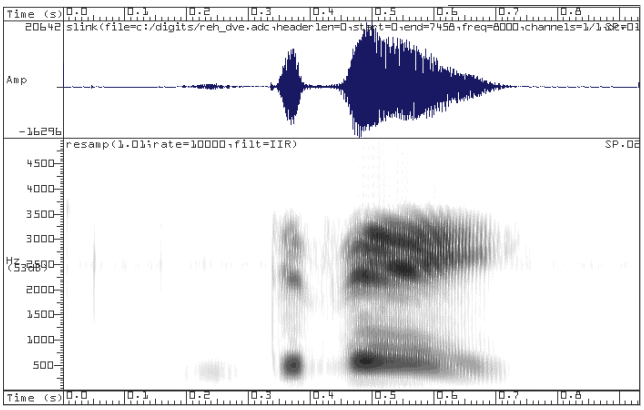

Figure 3. Example of the D2 dialectal pronunciation of Slovenian digit dve [SOUND 0340_03.wav].

The most frequent dialectal variations (D1 and D2) included the diphthongizations of /e/ in the S1 pronunciation of Slovenian digit dve. The D1 dialectal class could be characterized by the rules /e/ → /ɪə/ or /e/ → /ɪɛ/. For example, /ɪə/ denotes a diphthong with /ɪ/ and /ə/ initial and final steady state target positions, respectively. 9% of pronunciations of dve belonged to the D1 dialectal class. Example pronunciation of this dialectal variant is illustrated in Figure 2.

The second most frequent dialectal variant of the Slovenian digit dve, denoted by D2, could be described by the rules /e/ → /əɪ/ or /e/ → /ɛɪ/. As before, /əɪ/ denotes a diphthong with the /ə/ and /ɪ/ initial and final steady state target positions, respectively. The D2 class exhibited 6% of pronunciations in the database. Example pronunciation of the D2 variant is given in Figure 3.

Common to the D1 and D2 dialectal classes is the diphthongization of /e/ defined in the Standard Slovenian pronunciation. The main difference between them is that the initial and final steady state target positions of diphthongs are reversed, e. g., /ə/ and /ɪ/ initial and final steady state target positions of the D1 become /ɪ/ and /ə/ for the D2, respectively. Similar observation is valid for the other diphthongs in rules listed above. Another less frequent dialectal pronunciation of Slovenian digit dve could be characterized by the rule /e/ → /ɛ/, i.e., the close-mid unrounded front vowel /e/ is replaced by the open-mid unrounded front vowel /ɛ/. This dialectal class is denoted as D3 (Table 1) and exhibited 4% pronunciations in the database under study.

Pronunciation variants of the Slovenian digit pet `five´ are given in Table 2. Perceptual and instrumental analyses identified two dialectal pronunciation classes, D1 and D2 that differed from the Standard Slovenian pronunciation denoted as S /p e t/.

Pronunciation type	Pronunciations
S	74%
D1	22%
D2	4%

Table 2. Pronunciation variations of Slovenian digit pet `five´.

Table 2 shows that 74% pronunciations of digit pet in the database were classified as the Standard Slovenian. Example of the Standard Slovenian pronunciation of pet /p e t/ is given in Figure 4.

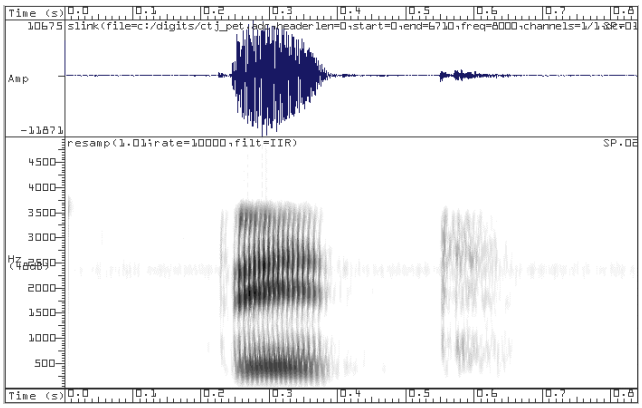

Figure 4. Example of the Standard Slovenian pronunciation (S) of digit pet `five´ [SOUND 0340_04.wav].

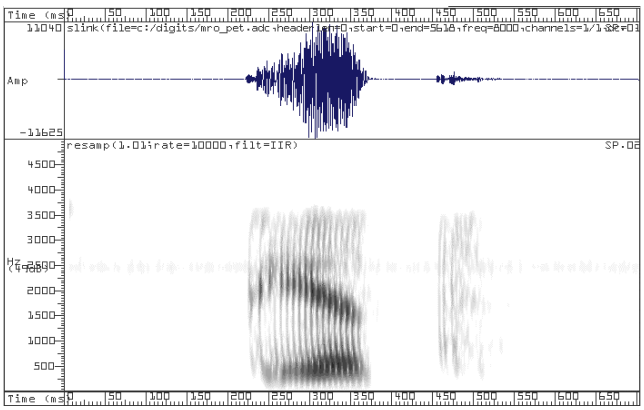

Figure 5. Example of the D1 dialectal pronunciation of Slovenian digit pet `five´ [SOUND 0340_05.wav].

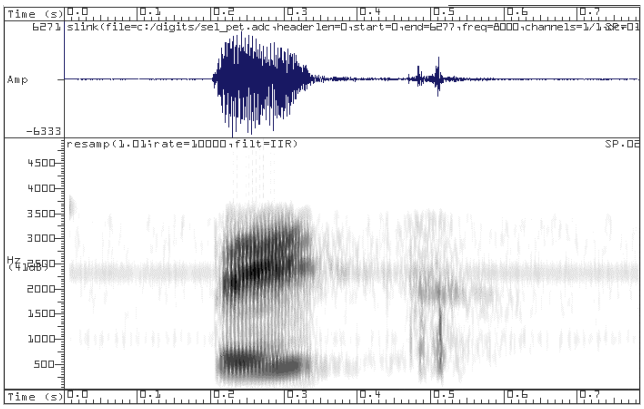

Figure 6. Example of the D2 dialectal pronunciation of Slovenian digit pet `five´ [SOUND 0340_06.wav]

The most frequent variation from the Standard Slovenian pronunciation of pet was the dialectal class D1. It included 22% pronunciations. As in the previous case of digit dve, the D1 dialectal class could be characterized by the rules /e/ → /ɪə/ or /e/ → /ɪɛ/. Therefore, /ɪə/ denotes the diphthong with /ɪ/ and /ə/ initial and final steady state target positions, respectively. Example pronunciation of the D1 variant of the digit pet is given in Figure 5.

As in the previous case of digit dve, the dialectal class D2 of the digit pet was less frequent. Similarly, this class could be described by the rules /e/ → /əɪ/ or /e/ → /ɛɪ/. As already defined above, /əɪ/ denotes a diphthong with /ə/ and /ɪ/ initial and final steady state target positions, respectively. The D2 class exhibited 4% of pronunciations in the database. An example pronunciation of the D2 variant of Slovenian digit pet is given in Figure 6.

In contrast to the pronunciation patterns of Slovenian digit dve, the digit pet included only two dialectal variants, D1 and D2. Dialectal pronunciation rules replacing the vowel quality without diphthongization were not observed on the database under study.

Table 3 summarizes results for pronunciation variants of the Slovenian digit devet `nine´. In this case three dialectal classes were identified in addition to the Standard Slovenian pronunciation.

Pronunciation type	Pronunciations
S	84%
D1	11%
D2	3%
D3	2%

Table 3. Pronunciation variations of the Slovenian digit ´devet´ (nine).

Example pronunciation of the Standard Slovenian digit devet is given in Figure 7. This type of pronunciation was identified in 84% on the database under study. Dialectal pronunciations were identified in 11% for the D1, 3% for D2, and 2% for D3.

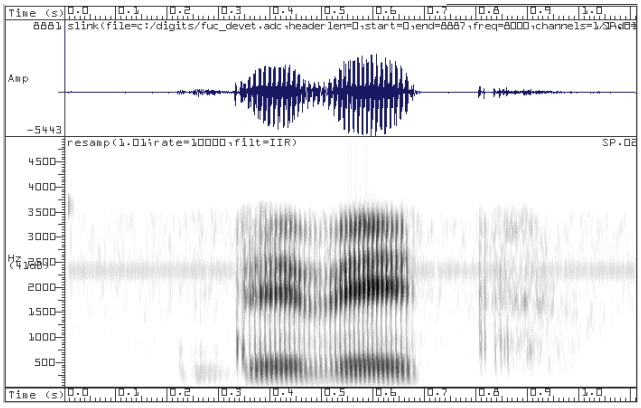

Figure 7. Example of the Standard Slovenian pronunciation of digit devet /d ɛ v e t/ `nine´ [SOUND 0340_07.wav].

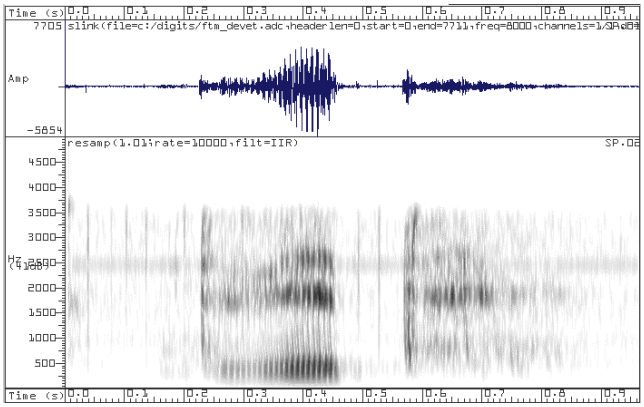

Figure 8. Example of the D3 dialectal pronunciation of Slovenian digit devet `nine´ [SOUND 0340_08.wav].

Dialectal classes D1 and D2 are defined by identical rules as in the case of digits dve and pet. Dialectal class D3, however, could be described by a rule /ɛ/ → /ə/. In many occasions the /ə/ was very short or omitted. Example of the D3 pronunciation of digit devet is illustrated in Figure 8.

3.3. Formant dynamics

Contrastive analysis between the standard and dialectal D1 pronunciations of Slovenian digit pet `five´ is given in Figure 9. For this purpose the formant contours were sampled at eight equidistant points (1: start of the contour, 8: end of the contour). Results for the D1 dialectal pronunciation exhibit clear diphthongization of /e/.

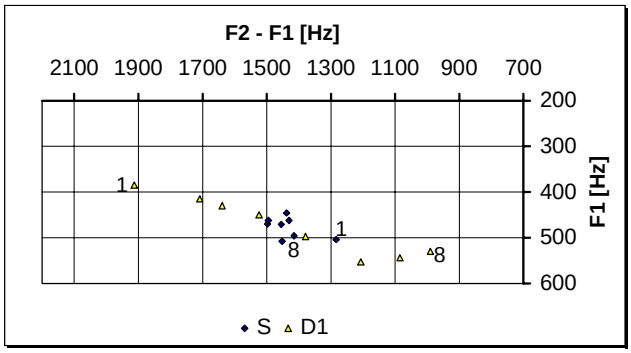

Figure 9. Formant dynamics for the Standard Slovenian (S) and dialectal (D1) pronunciations of digit pet `five´.

4. CONCLUSIONS

Slovenian language has seven major dialect regions and more than 40 dialects [1]. This presents a clear challenge for any mature interactive voice services system for Slovenian [2]. This work, however, has not addressed the dialectal pronunciations of Slovenian digits in each of the dialect regions or dialects. It attempts to answer the question: When Slovenian speaker is prompted in Standard Slovenian to pronounce a digit (0-9) which pronunciation variants could be expected and how could they be modeled for the robust ASR?

For this purpose the database of 300 (out of 780) speakers was investigated from the automatic recognition and acoustic phonetic points of view. The ASR experiments suggested that the most confusable digit pairs include the digits {/p e t/ pet (5), /d ɛ v e t/ devet (9)}, {/d v e/ dve (2), devet (9)} and {tri /t r ɪ/ (3), devet (9)} [5].

For the Slovenian digits dve, pet, devet further perceptual and instrumental analyses suggest that the number of pronunciation variants is less than the number of dialect regions. Furthermore, the degree of dialectal influence in digit pronunciation is not the same for each of the digits under study, i.e., the Standard pronunciations were identified in 81%, 74% and 84% for the digits dve, pet and devet, respectively. One would expect that the same speaker imposed the dialectal structure to each digit pronunciation. This was not the case in present study since the digit five /p e t/ exhibited a greater number of dialectal pronunciations than the other two digits.

Two the most frequent broad dialectal rules (D1 and D2) that could be used in refined digit modeling for the ASR were identified. They specify that a vowel in the Standard Slovenian pronunciation changes to a diphthong. The rule D1 applied more frequently than the rule D2 for all digits in the database under consideration. Formant dynamics analysis results support the view of diphthongization in dialectal pronunciations.

The above analysis could be performed on a bigger standarized database like the ELRA's Slovenian SpeechDat(II). This could also validate if the present database is representative enough to describe the dialectal structure of the Slovenian digit pronunciations over the fixed network.

ACKNOWLEDGMENTS

ASR experiments [5] were performed by the HTK [Young & the CUED speech group]. Main tool for acoustic phonetic analyses was the SFS [Huckvale, UCL].

REFERENCES

[1] Logar, T. 1993. Slovenian Dialects. Ljubljana: Mladinska knjiga.

[2] Petek, B. 1998. Slovenian Language in the Information Age. Proc. Foundation of Endangered Languages Conference, Edinburgh, 87-90.

[3] Petek, B. and Sustarsic, R. 1997. A Corpus-Based Approach to Diphthong Analysis of Standard Slovenian. Proc. ESCA Eurospeech'97, Rhodes, 767-770.

[4] Petek, B, Sustarsic, R., and Komar, S. 1996. An Acoustic Analysis of Contemporary Vowels of the Standard Slovenian Language. Proc. Int. Conf. on Spoken Language Processing ICSLP'96, Philadelphia, 133-136.

[5] Petek, B. 1995. Towards Voice-Interactive Telephone Services in Slovenia: On Prosody of Digits Using the Sociolinguistic Framework. Proc. ESCA Eurospeech'95, Madrid, 1015-1018.

[6] Sustarsic, R., Komar, S. and Petek, B. 1995. Illustrations of the IPA: Slovene. *Journal of the International Phonetic Association*, 25(2), 86-90.