IDENTIFICATION OF REGIONAL VARIANTS OF HIGH GERMAN FROM DIGIT SEQUENCES IN GERMAN TELEPHONE SPEECH

Christoph Draxler and Susanne Burger
Institut für Phonetik und Sprachliche Kommunikation
der Universität München
Schellingstr. 3, D-80799 Munich, Germany
Tel. +49 89 2180 {2807|2806}, Fax +49 89 2178 2652, E-mail: {draxler|burger}@phonetik.uni-muenchen.de

ABSTRACT

From the German SpeechDat(M) database of telephone speech the digit sequences items that were spoken as chains of individual digits were extracted. From these digit strings, a subset of 39 strings was selected by dialect experts and according to the region information provided by the speaker. The German federal states were used as region classes because this information can easily be provided by the speaker. 7 test persons were asked to listen to the subset of digit strings and to classify them by region. It was found that the overall success rate for the classification is 40%; if the regions neighboring the correct region are also counted as correct, the success rate is 68%.

1. INTRODUCTION

In the SpeechDat(M) [1] project, telephone speech (8 KHz, 8 bit alaw) was collected in 8 major EU languages: Danish, English, French, German, Italian, Portuguese, Spanish, and Swiss French. The goal of this data collection is to provide a common basis for the development of telephony applications and services, and to serve as a reference corpus for research in phonetics, phonology, and linguistics.

The recorded speech consists of expressions useful for voice-driven teleservices and applications (date and time, money amounts, application words and phrases, spelling, yes/no responses, phonetically rich sentences, and digit sequences). The databases contain 1000 speakers with a good geographical coverage and a balanced gender and age distribution.

In many telephony applications, digits and digit strings play a key role. The are used for identification purposes, e.g. credit card or account numbers, to indicate date and time, to denote quantities, e.g. money amounts, and for the selection of services.

In this paper, the digit sequence recordings of the German SpeechDat(M) corpus are used to examine whether such utterances are sufficient to extract region information from speakers. For telephony applications, this information may contribute to improve both speaker identification and speech recognition tasks; furthermore, it is of general interest to document the pronunciation of digits and digit strings across the regions of Germany.

2. EXPERIMENTS

An experiment was set up to determine whether digit strings spoken in German over the telephone are sufficient to determine the regional variant of the speaker's speech.

2.1 Data

In the SpeechDat project, the age at which speakers entered school was taken as decisive for the speaker's dialect. In the German SpeechDat(M) recordings, speakers were asked for the federal state of Germany where they entered primary school.

The SpeechDat(M) German database contains two digit sequence items, one with 16 digits grouped in fours, the other with 14 digits grouped in pairs. From the 2 * 1000 recordings of the digit sequences, 6 recordings were empty. From the remaining 1994 recordings, 454 were selected according to the following criteria: the transcription contains only isolated digit words, not number words. It does not contain mispronunciations or word fragments, nor noise markers, nor signal truncation markers. The regional coverage for the selected recordings matches that of the SpeechDat(M) database as a whole, i.e. it is imbalanced: southern Germany is over-represented, and eastern Germany is under-represented.

Experiment 1

Two experts classified the selected recordings. They could rate their selections on a scale of confidence with 4 values. The German federal states were used as a basis to define the region classes, with small or indistinguishable regions merged into larger ones, resulting in 12 classes. For example, HB was merged with NI, BE with BB, etc. The experts then selected a subset of 39 recordings for the actual experiment. This subset contains 20 male and 19 female speakers, and covers 11 of the 12 region classes.

For the experts, the main criteria for the classification were voiced/voiceless /s/ and differences in the quality of vowels and diphthongs (e.g. voiceless /s/ = BY, BW or SN) [2]. Both experts come from BY.

Experiment 2
7 test persons besides the experts were asked to perform the experiment, 3 male and 4 female. The regional origin and dialect was known for the test persons.
For this experiment, one of the experts chose a subset of two female/two male sequences for each region (where possible). Criteria for the selection were the classification of both experts and the sound quality of the signal files.

2.2 Technical setup

The region identification task was implemented using a simple WWW based tool derived from the WWWTranscribe system used for the SpeechDat transcriptions [3].
The main window consists of a simple form that contains an output button, a popup menu for the selection of a region, radio buttons for the confidence rating, and a save button to save the classification and proceed to the next recording (Fig. 1).

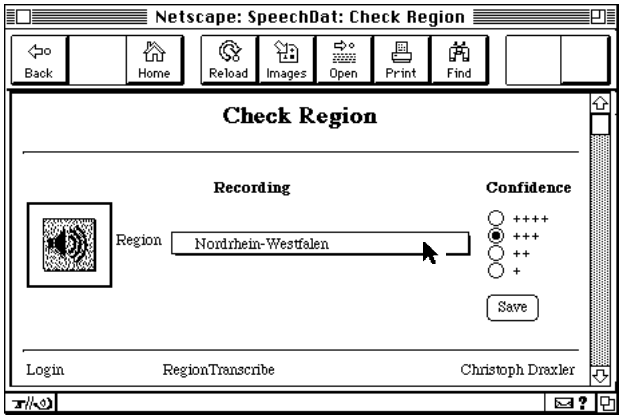

Fig. 1: main window screen shot

Each recording can be repeated as often as wanted. Once a recording had been classified, the test person cannot change the classification. All classifications are stored in a general log file.

3. Results

3.1 Results of experiment 1.

Table 1 shows the absolute number of the sequences of each region, the percentage of the correct decisions, the percentage for a broader analysis, the percentage of completely wrong decisions and the combined values for both experts. In the broader analysis, a decision was counted as correct if either the correct region or one of its immediate neighbors was identified (e.g. BY+BW+HE+SN+TH). Completely wrong decisions are those where not even a neighboring region was selected.
3.2 Discussion of experiment 1

The expert values shows that the female expert (s) performed better than the male (f) for correct decisions and broad analysis. Both experts has the highest correct identification rate for BY and BW, followed by SN.
The reason for this could be that both experts come from BY, BW is a neighbor of BY, both regions are the southern-most German regions. Furthermore, together with SN, they have dialects which are very different from that of other parts of Germany for digit sequences. Digit sequences from RP, HE, TH and ST were not or only badly recognized. RP and HE do not seem to have a well identifiable pronunciation for digits. Both regions are in the middle of Germany and extend into southern Germany. Southern HE and RP dialects are very similar to their southern neighbors, those of the northern parts of this regions sound more like the dialects of southern North Germany. The low identification rate of speakers from BB, TH and ST can be put down to the small number of speakers from these regions of the Speech-Dat(M) database. Another reason could be that „Sächsisch“, the dialect spoken in the East German region SN, was considered to be the prototypical pronunciation by West Germans to such a large extent that the existence of other regional variants was neglected.

3.3. Results of experiment 2

Table 2 shows the correct identification in percent for the digit sequences for each region and for each test person and the total identification rates. Results for the female test persons are printed in italics.
Table 3 is a confusion matrix for the regions with the true regions in the first column, the identified regions in the right columns. The values are computed for all test persons and are given in percent.
The confusion matrix in table 4 contains an overall comparison between the female and the male test persons for the regions that were identified either badly or not at all. Values are given in percent.
Table 5 displays a broad analysis of the results. Here a region is counted as correct even if it is only one of the neighbors of each region.

3.4. Discussion of experiment 2

The results are a similar to the results of the experts: a high identification rate for BY, BW and SN.
Just like in the expert analysis, the recognition of sequences from RP, HE, TH, ST and BB is very low (<20%). The reasons may be the same as for the results of the first experiment. Only in case of HE, NI and SH the female test persons have better identification rate than the male, but in total both genders have nearly the same rate.
An interesting point is that only the BY test persons identified digit sequences of their home region at 100%. The test persons from NI recognized the speakers from NI at 33%, HE recognized HE at 25% and NW recognized NW at 50%.

region class		count	expert s			expert f			total		
			correct	broad	wrong	correct	broad	wrong	correct	broad	wrong
Bayern	BY	**179**	73	93	7	53	86	14	63	90	10
Baden-Württemberg	BW	**41**	66	85	15	63	90	10	65	88	12
Rheinland-Pfalz	RP	**18**		44	56		11	89		28	72
Hessen	HE	**13**	23	77	23		85	15	12	81	19
Nordrhein-Westfalen	NW	**70**	30	74	26	26	79	21	28	76	24
Sachsen	SN	**12**	58	100		42	50	50	50	75	25
Thüringen	TH	**2**			100		100			50	50
Sachsen-Anhalt	ST	**4**			100		25	75		13	88
Brandenburg	BB	**17**	18	41	59		29	71	9	35	65
Niedersachsen	NI	**39**	23	79	21	41	82	18	32	81	19
Schleswig-Holstein	SH	**37**	19	30	70	8	30	70	14	30	70

Table 1: Identification rates by expert (in percent)

region	***NI_anj***	***NI_ing***	**NI_dra**	***BY_dan***	***BY_su***	**BY_kal**	**BY_fel**	**NW_sch**	***HE_ino***	***female***	**male**	**total**
BY	*50*	*75*	75	*100*	*100*	100	100	100	*75*	*80*	94	86
BW	*100*	*75*	100	*75*	*100*	100	100	75	*25*	*75*	94	83
RP		*25*	25							*5*	6	6
HE	*25*	*25*	50		*25*				*25*	*20*	13	17
NW	*25*		25	*25*	*100*	25	100	50	*50*	*40*	50	44
SN	*75*	*50*	50	*25*	*100*	75	75	75	*75*	*65*	69	67
TH												
ST												
BB				*25*	*25*	50				*10*	13	11
NI	*25*	*75*		*50*	*50*	25		25	*50*	*50*	13	33
SH	*50*	*50*	50	*50*	*75*	25	25	50	*25*	*50*	38	44
total	***36***	***38***	**38**	***36***	***59***	**41**	**41**	**38**	***33***	***41***	**40**	**40**

Table 2: Identification rates by region class and test person (in percent, *female* in italics)

Table 3 displays in the diagonal fields the percentage of correctly identified regions, and in the other fields the actual distribution of the incorrect identifications.

Especially those values are interesting where the correct region has no or only a small border in common with the identified region. RP has in most cases been mistaken for BY and NI, the regions HE, RP and TH were mistaken for NI, and ST and BB were mistaken for NW. Most of the incorrect identifications went to NI. The reason could be that inhabitants of NI are said to be nearly without dialect, therefore it can be assumed that NI served as a kind of trashbox for the test persons.

The comparison of the distribution of values between female and male for the badly identified regions shows interesting differences between female and male. HE is taken for a northern region by female test persons and as a southern region by male. TH is considered to be a western region by female test persons and close to the center by male.

Even if the neighbor regions are taken in account, RP has the lowest identification, followed by BB.

4. CONCLUSION

The results of the experiments must be seen as preliminary only for the following reasons:

The number of recordings is small. For some regions there were only very few suitable recordings, or even none at all (e.g. Mecklenburg-Vorpommern).

The material is not well balanced in terms of the prominence of regional speech phenomena. The high number of recordings for BY and BW allowed the selection of such recordings where the regional variation of German was really prominent – this was not possible for other regions.

With larger speech databases, e.g. SpeechDat(II), which for German will contain 4000 speakers with a well-balanced geographical distribution, these two problems can be overcome.

Other factors that contribute to the low rates of identification are:

In a formal situation, as in the SpeechDat dialogues, speakers tend to use high German, especially for read speech. This is particularly true for digits, because speakers try to articulate as clearly as possible to avoid misunderstandings.

The federal states do not match dialect regions very well. Some states have more than one clearly distinguishable dialect region. Other states are smaller than a language region; these states can thus be considered as belonging to neighbouring or surrounding states (as has been done here).

The main advantage of using federal states as classes is that they can be provided easily by the speaker and that

	BY	**BW**	**RP**	**HE**	**NW**	**SN**	**TH**	**ST**	**BB**	**NI**	**SH**	**count**
BY	86						3		3		3	**4**
BW		83	6	8				3				**4**
RP	33	11	6	14	11					22	3	**4**
HE	17		3	17	19			3		28	11	**4**
NW	6			6	44				3	28	11	**4**
SN	6	6	3		3	67	8	8				**4**
TH				22	22				11	33	11	**1**
ST	6			11	22		6		17	33	6	**2**
BB	3	8	17	8	22	6	3	6	11	17		**4**
NI	6	6	14	6	17			3		33	14	**4**
SH	3		8	6	11		3	3		22	44	**4**
total	***17***	***12***	***6***	***8***	***15***	***7***	***2***	***3***	***3***	***18***	***9***	***39***

Table 3: Confusion matrix by region (in percent)

		BY	**BW**	**RP**	**HE**	**NW**	**SN**	**TH**	**ST**	**BB**	**NI**	**SH**	**count**
BB	female	5	5	20	5	15	5	5	10	10	20		4
	male		13	13	13	31	6			13	13		4
HE	female			5	20	30					35	10	4
	male	38			13	6			6		19	13	4
RP	female	25	20	5	10	10					25	5	4
	male	44		6	19	13					19		4
ST	female				10	20		10		20	40		2
	male	13			13	25				13	25	13	2
TH	female					40				20	20	20	1
	male				50						50		1

Table 4: Confusion matrix by gender of the test person and region for wrong classifications (in percent)

region	***NI_anj***	***NI_ing***	**NI_dra**	***BY_dan***	***BY_su***	**BY_kal**	**BY_fel**	**NW_sch**	***HE_ino***	**total**
BY	*50*	*75*	75	*100*	*100*	100	100	100	*100*	89
BW	*100*	*100*	100	*100*	*100*	100	100	100	*75*	97
RP	*25*	*25*	50	*25*	*25*	0	25	75	*25*	31
HE	*75*	*100*	75	*100*	*100*	75	50	100	*75*	83
NW	*75*	*75*	50	*100*	*100*	25	100	75	*100*	78
SN	*75*	*75*	75	*100*	*100*	100	100	75	*100*	89
TH	*0*	*100*	100	*0*	*0*	100	100	100	*0*	56
ST	*100*	*100*	100	*100*	*0*	0	50	0	*50*	56
BB	*50*	*50*	25	*50*	*50*	75	25	0	*25*	39
NI	*100*	*75*	50	*75*	*50*	50	75	75	*100*	72
SH	*50*	*75*	25	*75*	*50*	50	0	25	*75*	47
total	***67***	***74***	**62**	***79***	***69***	**62**	**64**	**67**	***72***	**68**

Table 5: Broad identification rates by region class and test person (in percent, *female* in italics)

there are only 16 states. An alternative could be to ask for the ZIP code of the town the speaker entered school and to base the region classes on the first two digits of this ZIP code. However, this information will be unreliable because a) at the time older speakers entered schools ZIP codes were not in use, and b) ZIP codes in Germany were changed in 1993.

We plan to repeat this experiment for the larger SpeechDat(II) database, and also for the recordings of the AT+T project currently being recorded at the Phonetics Department. In this project, the true dialect region is known for each speaker.

5. REFERENCES

[1] SpeechDat(M): EU-project LRE-63314

[2] König, W.; dtv-Atlas zur deutschen Sprache, dtv-Verlag, München, 1978

[3] Draxler, Chr.; WWWTranscribe - A Modular Transcription System Based on the World Wide Web; Eurospeech ‘97, Rhodos, 1997