**In:** Hammwöhner, Rainer; Wolff, Christian; Womser-Hacker, Christa (Hg.): Information und Mobilität, Optimierung und Vermeidung von Mobilität durch Information. Proceedings des 8. Internationalen Symposiums für Informationswissenschaft (ISI 2002), Regensburg, 8. – 11. Oktober 2002. Konstanz: UVK Verlagsgesellschaft mbH, 2002. S. 21 – 25

# Effects of Manual Phonetic Transcriptions on Recognition Accuracy of Streetnames

*Thomas Bauer*[1], *Ludwig Hitzenberger*[1], *Marcus Hennecke*[2]

[1]Information Science
University of Regensburg
D-93040 Regensburg
Germany
{Thomas.Bauer, Ludwig.Hitzenberger}
@sprachlit.uni-regensburg.de

[2]Temic Speech Processing
Söflinger Str. 100
D-89077 Ulm
Germany
marcus.hennecke@temic.com

**Abstract**
In automotive systems, the accurate recognition of street- and place names as needed e.g. for speech controlled navigation systems is difficult, since many of them do not follow regular transcription rules. We report on experiments investigating the effects of correct manual transcription versus strictly automatic transcription, with regard to recognition accuracy. Our experiments showed that in case of large recognizer lexica, the recognition rate could be increased by about 10% by using a manually corrected versions of the phonetic transcriptions used for the recognition process.

## 1 Introduction

In automotive systems, speech recognition and dialogue systems are becoming more and more important. At the same time navigation systems belong to the standard equipment in cars and, obviously used while driving, speech seems to be the only adequate means of interaction[1]. These systems need to handle an enormous amount of data. There are about 60000 place names and more than half a million street names and points of interest in Germany. Proper names are typically a problem in speech -recognition [4], especially if automatic transcription is used since they often do not follow regular transcription -rules. Streetnames are even worse, with many of them combining words from different languages. In Germany we often see names like John-F.-Kennedy-Platz or Clermont-Ferrand-Allee. Another problem is the lack of

context for the disambiguation which we normally encounter in other dialogue -system domains. There is also a very flat hierarchy of the search space (e.g. the lowest level of the street name list is the concerned city). Thus, the recognition rate on the word level is crucial for the success of the dialogue. One way to improve the recognition rate is supposed to be the adequate phonetic transcription of names with respect to the monolingual recognition system. The transcription of the vocabulary items typically follows static rules applied through an automatic transcription device. On the other hand manual transcription is very costly and time consuming, but so far the only way to achieve more correct transcriptions and thus better recognition rates. In this paper we report on experiments designed to measure the achievable improvements using manual transcriptions over automatic transcriptions.

# 2 Experimental setup

## 2.1 Transcriptions

For our experiments we used a data set consisting of the street names of five big German cities (Munich, Stuttgart, Sindelfingen, Ingolstadt, and Regensburg) with a total of 11437 street names in the official orthographic representation taken from a standard navigation database provided by Navigation Technologies (NavTech)[1]. These items were transcribed automatically using the internal automatic transcription of the lexicon tool of the Temic speech-recognizer [2]. The output was subsequently checked by a phonetic expert and corrected if necessary. This correction was supported by an automatic transcription tool that uses a learning algorithm implemented by Sympalog[2]. About 15% of the data had to be corrected manually. The system, trained by these corrections, subsequently changed another 50% of the automatic transcriptions. The transcription alphabet used was a slightly modified SAMPA representation.

[1] NavTech is a company providing global geographical navigation data. Further information can be found at www.navtech.com.

[2] Sympalog is a company providing dialogue-systems and tools. Further information can be found at www.sympalog.com.

### 2.2 Vocabularies

For the tests we produced four different sets of vocabulary using the lexicon tool of the Temic recognizer. Thus we have two independent variables, the vocabulary size and the correctness of the phonetic transcriptions.

| Name | Number of items | Description |
|---|---|---|
| Vok1 | 11437 | Automatic transcription without manual corrections. |
| Vok2 | 11437 | Automatic transcription including manual corrections. |
| Vok1_s | 110 | Automatic transcription without manual corrections. |
| Vok2_s | 110 | Automatic transcription including manual corrections. |

Table 1: Vocabularies

### 2.3 Test data

A random set of 110 street names was spoken and recorded by 34 test persons. We had 17 female and 17 male speakers with an average age of 29 years. These test persons used a recording tool and recorded the names unsupervised. The audio format was 44100Hz/16Bit. Afterwards the recorded samples were converted to 8000Hz/16Bit PCM raw data (the proper format for the recognizer) using CoolEdit pro 1.1.

### 1.4 Speech recognition

The test data were processed by the Temic speech recognizer[3] [3] on a WindowsNT 4.0 platform using the different vocabularies (Vok1, Vok2, Vok1_s, Vok2_s). The recognizer was activated by Microsoft Access using VBA. The recognition results were stored in this Access database.

## 2 Results

### 2.1 Recognition accuracy

| | Vok1 | Vok2 | Vok1_s | Vok2_s |
|---|---|---|---|---|
| No. of samples tested | 3740 | 3740 | 3740 | 3740 |
| No. of samples recognized correctly | 2763 | 3127 | 3374 | 3600 |
| Recognition Rate [%] | 73.9 | 83.6 | 90.2 | 96.3 |

[3] Temic Star Rec DSR 1.7.

Table 2: Recognition rates

The table 2 shows the recognition rates of the four different vocabularies. The number of samples is the total number of all utterances in the test data. Since each test person recorded all the street names, this sums up to 3740 samples. The following box plot shows the distribution of the recognized items and the appropriate confidence interval (95%). The grey box represents the second and third quartile of the data, while the line in the box indicates the median of the data.

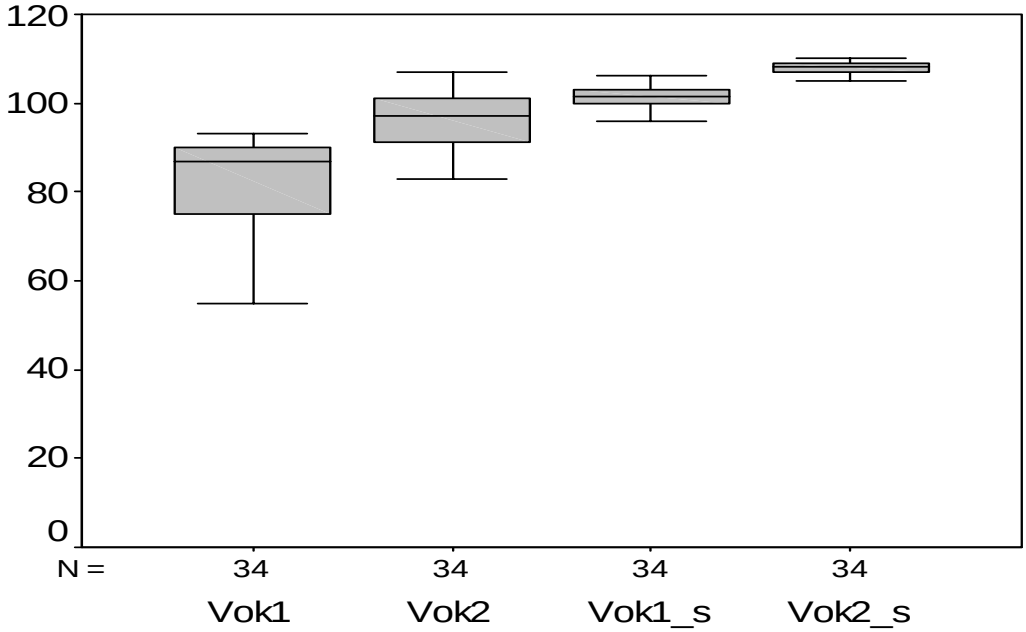


Diagram 1: Distribution of absolute number of correctly recognized items and corresponding vocabulary

## 2.2 Significance

As table 3 shows, all of the tests were highly significant.

| Pairs | Mean of the difference | Standard deviation | Standard error of the mean | 95% confidence interval of the difference | | Significance |
|---|---|---|---|---|---|---|
| | | | | lower | upper | |
| Vok1 Vok2 | -10.71 | 3.52 | .60 | -11.93 | -9.48 | .000 |
| Vok1 Vok1_s | -17.97 | 9.19 | 1.58 | -21.18 | -14.77 | .000 |
| Vok1 Vok2_s | -24.62 | 9.58 | 1.64 | -27.96 | -21.28 | .000 |
| Vok2 Vok1_s | -7.26 | 10.13 | 1.74 | -10.80 | -3.73 | .000 |
| Vok2 Vok2_s | -13.91 | 10.01 | 1.72 | -17.40 | -10.42 | .000 |
| Vok1_s Vok2_s | -6.65 | 2.56 | .44 | -7.54 | -5.75 | .000 |

Table 3: T-Test of paired samples

# 3 Discussion

The recognition results show a very strong effect of correct transcriptions in the vocabulary of the recognizer. For the larger vocabulary we achieved an improvement from 73.9% to 83.6. This is a difference of 9.7%. For the small vocabulary there was an improvement from 90.2 to 96.3%, which is a difference of 6.1%. This indicates that the effect is stronger on larger vocabularies; which will be subject of further investigations.

# 4 Conclusions

Although it is very time consuming and costly to manually correct the transcriptions of large vocabularies, in the context of navigation data and proper names like street and place names there is an obvious benefit in doing so. This is exactly what our experiments show. Also the learning transcription algorithms benefit from the correction of the data, so that the intellectual effort will gradually diminish, but a certain amount of manual control will always be necessary.

# 5 References

Hitzenberger, Ludwig: Man Machine Interaction in Car Information Systems, LREC´98, Vol.1:179-182, 1998.

Lehtinen, G. et. al. IDAS: Interactive Directory Assistance Services, Voice Operated Telecom Services 2000, Ghent.

StarRec™ GDS (Grammar Development System). TEMIC Telefunken microelectronic GmbH, Ulm, Germany, 2000.

StarRec™ DSR (DaimlerChrysler Speech Recogniser). TEMIC Telefunken microelectronic GmbH, Ulm, Germany, 2000.

Lehtinen, G. et. al. IDAS: Interactive Directory Assistance Services, Voice Operated Telecom Services 2000, Ghent.