

ON USABILITY OF NAME DIALLING

Cristina Dobrin, Péter Boda, Kari Laurila

Nokia Research Center, Speech and Audio Systems Laboratory, Tampere/Helsinki, Finland
E-mail: {cristina.dobrin, peter.boda, kari.laurila}@nokia.com

ABSTRACT

Name dialling, one of the telephony speech recognition applications that became recently widely available, can be implemented using either speaker-dependent or speaker-independent speech recognition technology. In this paper we compare two experimental name dialling systems, representing both technologies, with respect to performance, usage and subjective users' opinion. Both systems are implemented as telephone network services and they are offered for continuous usage to the personnel of the authors' laboratory. The first system, named MATTI, relies on user-trained name tags and utilises speaker-dependent whole-word recognition technology. The other system, TEPPPO, offers a ready-made list of names using speaker-independent sub-word modelling approach. Despite superior recognition accuracy of the speaker-dependent recognition engine, experimental results show that the average real world success rate achieved with TEPPPO is higher (84% compared to 70%), learning is faster, user feedback is more positive and the system is more eagerly used by the personnel.

1. INTRODUCTION

During the past several years significant progress has been made in the field of automatic speech recognition (ASR). The speech recognition technology has advanced to a level where deployments in consumer products and telephone services have become reality. One important application, name dialling, offers a fast and easy way to set up calls, with no need to remember telephone numbers. While name dialling is simple from interaction point of view (one definite user goal, small number of possible paths in the interaction flow), it is not such from the usability point of view. Issues like ease-of-use, learnability and user acceptance are still to be evaluated for various systems.

Recently, quite a number of deployed and experimental name dialling systems were presented in the literature. Most of these systems are using speaker-independent speech recognition technology offering directory assistance (call routing) services, [1][2][3]. Speaker-dependent systems, where the user is requested to build up his/her own directory of name tags, are presented rarely [4]. Though there are no standards or general methods for measuring adequacy of interactive speech-enabled applications [5], systems are mostly described with interaction success rate, interaction time, caller behaviour, system usage, etc.

Our goal with the work presented in this paper is to describe usability aspects of speaker-dependent and speaker-independent name dialling technology. Via quantitative and qualitative measurements of two experimental systems we established a

picture about users' preference for name dialling strategy. The experimental systems presented in this paper are named as MATTI (the speaker-dependent system) and TEPPPO (the speaker-independent system). Before taking the MATTI system into use, the user has to train personal name tags. A name tag can contain whatever the user desires: first name, full name, nick name or some other words identifying a person (e.g. "my brother"). The training is done from a single utterance, which is convenient for the user but more risky approach from the system performance point of view. If the user is not co-operative enough, the system can erroneously accept an utterance which is not appropriate and thus the resulting model is not representing properly the corresponding utterance. This potential problem is eliminated in the other system. TEPPPO users are offered a pre-defined list of names (like in a corporate dialling application). The addition of new names in TEPPPO is done via a web interface.

The paper is organised as follows. In Section 2 we discuss the two systems and their implementation. Section 3 describes the testing methodology and the aspects of the comparison. In Section 4 we present the results of our study. Finally, we summarise the analysis results and draw some conclusions.

2. DESCRIPTION OF THE SYSTEMS

Both systems were developed on a UNIX-based computer-telephony integration platform which supports the implementation and hosting of a wide range of interactive voice response applications. The applications were implemented using proprietary speech recognition engines.

The two systems have identical user interfaces in the dialling phase. The basic interaction between the system and the user is as follows:

System: "Name, please."
User: <NAME>
System: "Calling <NAME>."

If a substitution error occurs (a wrong name is recognised), the user can cancel the call by pressing any key. In case the recogniser fails to output a name, the system informs the user about the problem and asks the user to utter the name again.

The following features were common for both systems:

- *prompt* content and audio quality;
- *user groups* consisting of novice and experienced users;
- unrestricted accessing and storing of *name tags*;
- continuous *monitoring* over similar time spans.

The systems differed in their underlying ASR technology (speaker-dependent vs. speaker-independent), which caused differences in how name tags were added to the directories used by the systems.

2.1 Description of the MATTI system

In MATTI the speech recognition engine uses speaker-dependent whole-word HMMs, trained from a single utterance. Each user creates his/her own personal name directory by training name tags attached to the corresponding telephone numbers. There is no maximum limit how many name tags a user can train. No restriction is imposed on the users either what the name tags should contain: users are free to train given names, full names (given name and surname) or something else. The users are capable of initiating calls by saying one of the stored names and then the call is transferred to the number associated with the recognised name. The basic functionalities of MATTI are: calling, training and browsing. In the browsing mode the user can listen back to the trained name tags and attached telephone numbers, re-train name tags and delete name tags. Users were given a brief manual with instructions on how to access and use the system.

2.2 Description of the TEPPPO system

The TEPPPO system utilises speaker-independent context-independent phoneme models that were estimated using a phonetically rich speech database consisting recordings from 700 Finnish speakers. Models for each name were constructed by concatenating phoneme models according to the pronunciation rules. Foreign names were also modelled with Finnish phonemes, which somewhat deteriorated the recognition accuracy. The names of our laboratory personnel formed the base vocabulary of the system, containing more than 100 names. The ready-made name list consisted of full names, including the given name and the last name of each employee. TEPPPO users could immediately make calls to their colleagues without the need to train *à priori*. In order to offer individualised name lists, the users were able to add more names via a web interface. Browsing the name list was possible only via the web interface.

3. TESTING METHODOLOGY

The two name dialling systems provided a good test-bed to experiment with automatic speech recognition in real-world applications used on a daily basis. Both systems were introduced to the users by e-mail and presentations. In addition, detailed manuals were accessible to the users if they needed. The testing periods reported in this paper were about three months. MATTI was launched first and after several months later the TEPPPO system was released. Both systems had over 50 users, mainly native Finnish speakers. During the testing periods it was the choice of the users whether they made their phone calls in a conventional way or via the systems. The systems could be accessed from any phone: office, home or mobile, including hands-free car installations.

All interactions between the systems and users were logged, system performance and user's behaviour were monitored

continuously. The analysis of logged data was performed semi-automatically across users and over time. Subjective opinions were gathered via user questionnaires for both quantitative and qualitative analysis purposes. The users were interviewed about different functionalities and overall impressions of the systems after one month of usage of each system. The results and findings of these analyses were used to further improve the systems.

3.1 Comparison terms

The systems were compared using the following objective criteria: interaction success rate, system usage, frequency of name usage, addition/training of new names, interaction time and learnability. Subjective measures like ease-of-use, usefulness, and perceived accuracy were determined from the questionnaires.

Success rate is the basic measure of the goodness of the system and tells us how often the application fulfils its purpose. The success rate was calculated automatically from the number of calls the users made with the system. An attempt was considered successful if the user allowed the system to transfer the call to the telephone number associated with the recognised name. If the recognition failed, the user was asked to utter the name again and the attempt was classified as unsuccessful. The user also had the possibility to cancel the call set-up by pressing any key: if the user prevented the call set-up, the attempt was classified as unsuccessful.

The number of times the users access the system is also a relevant measure of the acceptability of a system. System usage was evaluated on a daily basis.

The feature of adding new names, either by training new name tags (for MATTI) or by typing new names into the personal name list via the web interface (for TEPPPO), was evaluated in terms of number and type of added names. Names were classified into two categories: full names and short names. A combination of a given and a surname was considered as a full name, a given name or a surname alone was considered as a short name.

4. RESULTS

Database results. In order to get an idea about the basic recognition rates of speaker-dependent (SD) and speaker-independent (SI) recognition engines, an off-line microphone (non-telephone) database experiment was conducted. Two different vocabularies were used: one containing 30 Finnish first names and the other containing 30 Finnish full names (first and last names). In the experiment the recognition accuracy was measured with the original quiet environment samples and with noise corrupted samples into which car noise was added with signal-to-noise ratios of 5 and 0 dB. Results in Table 1 show that the speaker-dependent recognition engine obtained higher recognition accuracy in all test cases. Consequently, higher recognition accuracy and thus higher success rate were naturally expected for the MATTI system.

	SD engine first names	SI engine first names	SD engine full names	SI engine full names
Quiet	100%	94%	100%	100%
+5 dB	98%	80%	100%	99%
0 dB	96%	69%	99%	95%

Table 1. Performances of speaker-dependent (SD) and speaker-independent (SI) engines using off-line database.

Number and type of trained/added names. The number of trained names in MATTI varies very much among users, as can be seen in Figure 1. Some users have only a couple of trained names, while others may have tens of them. The average number of name tags is 11. About 64% of these names are short names (one-word, containing either a given name or a surname). The average duration of name tags is 0.7 sec. In the TEPPPO system users added names textually via the web interface. This time, the number of added names was considerably smaller, in average only 5 names were added by each user. The names that were added were mostly long ones (over 93%), which can be attributed to the fact that the base list in TEPPPO consisted of full names of colleagues thus hinting the users to use similar name structures also for the added names. TEPPPO users added less names in average than MATTI users, 5 vs. 11, since they did not have to add their colleagues' name which were already included in TEPPPO's base list. In Figure 1 the amount of names trained/added in the two systems are shown.

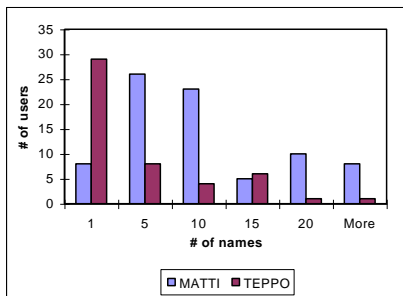


Figure 1. The amount of trained (MATTI) and added (TEPPPO) names.

Name usage frequency. In both MATTI and TEPPPO, users may train or add names that they never use actually. We looked into the log files and found out that the average number of intensively (more than ten times) used names in MATTI is 74% of the total amount of trained names. This represents about 8 names in average. For TEPPPO, this number is 10. The distribution of the most frequently used names is depicted in Figure 2.

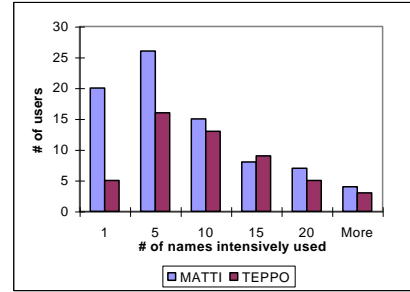


Figure 2. Frequency of usage of trained (MATTI) and added (TEPPPO) names.

The log information of TEPPPO revealed that about half of the intensively used names is coming from the readily offered base list, the other half from the user-added list. It is interesting to notice the “economical” behaviour of TEPPPO users: all the 5 user-added names (see above) are used intensively, in the average sense. Users thus added names to the base list only if they were sure that they really need these names.

Success rate. The average success rate obtained with MATTI was 70% while for TEPPPO it was 84%. The success rate for TEPPPO was higher despite the fact that the base name list contained over 100 names. The results contradict somewhat the database results, which indicated that the speaker-dependent models perform better. However, we explain this result by the fact that MATTI users preferred to train names that were rather short (e.g. “Ed” instead of “Eddie Jackson”). Also, the training algorithm had accepted training utterances that contained extra noises and some utterances that were falsely end-pointed, resulting in truncated or extended name tags. Moreover, MATTI users were speaking more out-of-vocabulary names, partially due to the fact that they did not remember how they had pronounced the names in the training phase (given name only, nickname or full name). The pre-existing name list in TEPPPO ensured the consistency and goodness of the models used for recognition. The TEPPPO base name list, as well as the user-defined names, included practically full names only which resulted in improved recognition accuracy.

Daily usage. The MATTI system was called daily by 16 users in average, around 35 times a day, and TEPPPO was called by 18 users in average, around 52 times per day. One user called MATTI in average 2.1 times a day and TEPPPO around 2.8 times a day, which shows that TEPPPO was more readily accepted by the users. The daily usage for the two systems is depicted in Figure 3. For viewing convenience, the two graphs have been smoothed and aligned. From Figure 3 it can be seen that the MATTI system usage decreased towards the end of the testing period (and was practically not used after the testing period), while the usage of TEPPPO remained high during the whole testing period and was/is frequently used even after that.

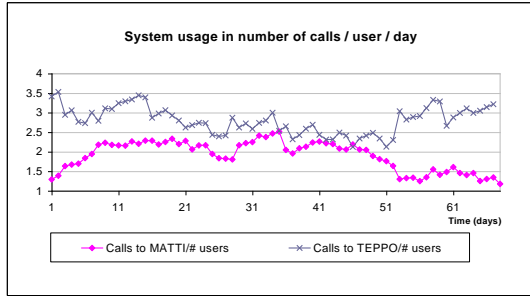


Figure 3. Comparison of daily system usage.

Interaction time. The system interaction time (measured from the moment the system answered the user's call until the end of the "Calling <NAME>." prompt) lasted for 12 sec in average, for both MATTI and TEPP0. Keying in a 10-digit telephone number takes about 15 sec, if the user remembers the number by heart. If not, because the user wants to call a colleague for instance, it takes about 25 seconds to open the Nokia corporate directory web page, typing the name of the colleague in, retrieving the number from the directory and keying it in. Compared to this traditional method, TEPP0 users save half of the interaction time.

Learnability. It is well known that the success rate varies according to the experience of users. We found out that the success rate after only 2 sessions is 54% for MATTI and 71% for TEPP0. These are important figures since the first impression about a system and the success achieved with it can essentially influence users' further behaviour. The learning effect is well observable for both systems: the average success rate for those who made more than 40 sessions arose to 79% in case of MATTI and to 91% with TEPP0. In order to evaluate how quickly the user learns to use the system, the term of objective learnability was defined as the average number of sessions after 90% of the final success rate is attained. For MATTI this parameter was 30 and for TEPP0 it was 21 sessions, which shows that it was faster to learn to use TEPP0. The average success rate is depicted along the number of sessions in Figure 4.

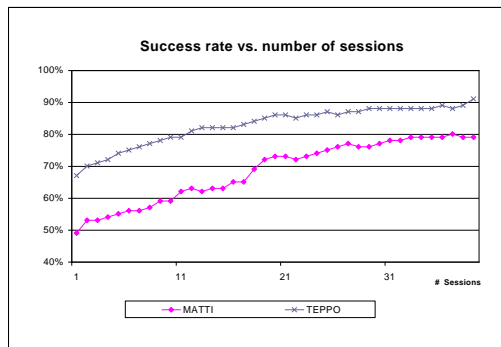


Figure 4. The effect of learning for the two systems.

Subjective opinions. Recognition accuracy and success rate, the traditional measures of automatic speech recognition and

speech-enabled applications, are helpful but do not necessarily provide direct insight into the overall system performance. Objective measures have to be analysed in conjunction with subjective opinions. Users' opinions provide valuable help to improve the system's ease-of-use. Our interview questionnaires focused on several issues, divided into perceived recognition accuracy, number of trained names, error correction and overall opinion.

Most of the users perceived the recognition accuracy as being very high, 81% and 83% of the users of MATTI and TEPP0, respectively, gave the highest grade. In average sense, the users said that they trained a slightly less amount of names than our results from objective measurements showed, indicating that the users had forgotten some of the trained names. Overall, both systems were perceived as easy-to-use, accurate, useful, fast in dialling mode and safe to use in a car. The biggest problem of MATTI was the slowness in the training phase. The TEPP0 system alleviated this problem by offering ready-made list of names.

Summary. The most important performance figures, grouped according to analysis aspects, are presented in Table 2. These results indicate that the TEPP0 system with speaker-independent speech recognition technology, readily compiled name list and possibility to add new names via a web interface offers a more attractive service than the speaker-dependent MATTI system, both in objective and subjective sense.

Comparison aspects	MATTI	TEPP0
Name Dialling		
Success rate (%)	70	84
Usage per day (calls/user/day)	2.1	2.8
Objective learnability (sessions)	30	21
Adding new names		
Number of new names	11	5
Type of names (short/long %)	64 / 36	7 / 93
Frequency of usage of names (%)	74	base list: 5 added: 100
Subjective opinion (better system)		
Perceived system accuracy		3
Adding new names		3
Usefulness		3

Table 2. Summary of the most important results.

5. CONCLUSIONS

In this paper two approaches in name dialling, namely speaker-dependent and speaker-independent technologies, were compared at the system level. User acceptance was evaluated using several parameters including success rate, system usage and objective learnability. Even though the simulation results indicated that our speaker-dependent recognition engine is better in terms of basic recognition accuracy, our experiments showed that a higher real world success rate was obtained with the system that utilised speaker-independent recognition engine (84% compared to 70%). The ready-made vocabulary used with the speaker-independent system ensured the consistency of the names used for recognition. Our analysis showed that the

speaker-independent system was more eagerly used (2.8 vs. 2.1 sessions/day/user) and the learning of its usage was faster. Regarding the subjective users' opinions, we found that success rates of about 70-80% were considered high among the users, and that the speaker-independent system was considered more useful, easier to learn and use.

6. REFERENCES

- [1] Fraser N. M., Salmon B. and Thomas T. "Call routing by name recognition: field trial results for the Operetta™ system". *Proceedings of IVTTA'96*, Basking Ridge, New Jersey, USA, 1996, pages 101-104.
- [2] Kellner A., Rueber B. and Seide F. "A voice-controlled automatic telephone switchboard and directory information system". *Proceedings of IVTTA'96*, Basking Ridge, New Jersey, USA, 1996, pages 117-120.
- [3] Billi R., Canavesio F. and Rullent C. "Automation of Telecom Italia directory assistance service: field trial results". *Proceedings of IVTTA'98*, Torino, Italy, 1998, pages 11-16.
- [4] Vysotsky G. J. "Progress in deployment and further development of the NYNEX VoiceDialing™ service". *Proceedings of IVTTA'96*, Basking Ridge, New Jersey, USA, 1996, pages 16-20.
- [5] Bernsen N. O., Dybkjær H. and Dybkjær L. *Designing Interactive Speech Systems: From First Ideas to User Testing*. Springer-Verlag, London, 1998.