Performance analysis of a multilingual directory enquiries application

Charl van Heerden, Marelie H Davel and Etienne Barnard
Multilingual Speech Technologies
North-West University, Vanderbijlpark, South Africa
Email: cvheerden@gmail.com

***Abstract*—In a multilingual society such as South Africa, a practical directory enquiries (DE) application should be able to serve users from various language backgrounds with information relating to names in various languages: a difficult task for automated speech recognition-based systems to perform. We describe the implementation of such a DE application (a 'Municipality Hotline') and evaluate its performance, focussing on its usability by callers from various linguistic backgrounds.**

I. Introduction

Directory enquiries (DE) have been an important commercial activity worldwide for several decades [1]. Initially, these services relied on human operators to elicit and recognize callers' requests, find the relevant information in a database, and speak back the information required (telephone number and/or address). Given the scale of DE services worldwide – for example, in the United States alone, over 6 billion such calls were made annually in the early 1990's [2] – it was clear that any form of automation would be economically important. Hence, a sequence of automated processes have become standard practice in DE applications during the past forty years, starting with a recorded prompt for request elicitation, and expanding to include automated readback of the requested information and compressed record-and-playback of caller requests to the operators.

Automation of the speech recognition process itself has become a practical reality in the past twenty years [2], as automatic speech recognition (ASR) accuracy and speed have improved to the extent that useful automation rates can be achieved. This is not to say that complete DE automation has become the norm: significant technical challenges imply that even the most sophisticated systems still rely strongly on operator assistance [1], and the lack of speech technology in most of the world's languages [3] implies that such capabilities are only available in wealthy developed countries such as the USA, Germany and the UK.

This need for localised technologies has been a driving factor behind the South African Directory Enquiries (SADE) project, which aimed to support ASR-based DE automation in the South African environment. Specifically, this project produced the necessary technologies to support an end-to-end ASR-based DE demonstrator application, able to provide a practical DE service in South Africa.

In this paper we evaluate the performance of the demonstrator application developed during the SADE project, focusing on its usability by system users. We first review some of the pertinent factors to consider when developing such an application (section II) before describing the specific choices made during the development of the current system (section III). The evaluation protocol is described in section IV and results are analysed in section V.

II. Background

The ASR task in DE is conceptually simple [4]: based on a spoken utterance, U, decide which of the directory listings L was most likely intended by a caller. However, this conceptual simplicity hides substantial practical complexities.

First and foremost, the range of possible utterances is extremely large in a typical DE task. The number of unique entries in a typical index may range from hundreds of thousands for a smallish local directory to tens of millions for a state-of-the-art North-American directory [5]. In addition, each of these entries may be requested in several ways: the caller may embed the request in a natural-language carrier ('I would like ...'), and may use various proxies that describe the requested entry equally well. Accommodating a large number of requests increases the difficulty of providing *natural* caller-system interaction, an important aspect of widely usable systems, as discussed further in Section II-A

Another important consideration relates to response times and interaction speeds. Human operators have improved significantly in the past decades: whereas the average work time for operators to complete a 411 call in the US was approximately 35 seconds in the late 1980s [6], work time per call had improved to less than 20 seconds by 2001 [7]. It may be acceptable for automated systems to be a little slower, since operator costs are a major driver for these speed improvements, but it is clear that caller satisfaction will be impacted if substantially longer interaction times are required.

Achieving these goals requires a careful integration of three major building blocks, namely the user interface, the speech-recognition module and the directory database.

A. User interface

One of the most important distinctions between user interfaces for DE systems rests on the amount of freedom the caller has to steer the interaction. At the one extreme, 'natural language' interfaces allow callers to phrase their requests using any wording they prefer; at the other extreme, 'directed dialogues' ask specific questions, guiding callers to provide a constrained set of answers. Between these opposites are 'mixed initiative' dialogues, in which the system generally guides the

interaction, but attempts to respond appropriately when the caller provides additional (but relevant) information. Directed dialogues were the first to be studied seriously [2], and remain the most widely popular style [8].

The success of Web search by voice has motivated system designers to pursue a more open-ended approach to user-interface design for DE [4], [5]. As we describe below, it is possible to create an efficient statistical language model that captures a wide variety of DE-related utterances; this model produces as output a recognition hypothesis (or a lattice of such hypotheses) which can then be matched against the directory to infer the desired listing. Thus, an open-ended prompt such as 'What listing are you searching for?' can be employed. However, the need for additional disambiguation is then increased; for Web search, a graphical user interface is generally available for the disambiguation process, but for DE, this remains a challenge [4].

Disambiguation, confirmation and error recovery strategies all impact on DE system performance [2], [8]:

- *Disambiguation* is required when the information provided by the caller is not sufficient to identify the intended listing uniquely (e.g. if there are several residential listings in the same city sharing the same initial and surname), or when the recognition output is likely to be confusable with an alternative entry (e.g. Barker and Parker).
- *Confirmation* generally translates to a simple trade-off between accuracy and speed: recognizers invariably provide a measure of confidence for the recognition result obtained, and when lower confidence thresholds are employed, users are asked to confirm their selections more frequently, but the likelihood of providing an incorrect answer is also reduced.
- The basic choice in *error recovery* is to either repeat the prompt (or prompt sequence) that lead to the initial error - in the hope that the later utterances will be recognized more accurately - or to employ alternative dialogue strategies in order to improve the chances for success.

B. Speech recognition

The ASR system must be highly accurate in recognizing caller utterances, but also provide reliable confidence estimates, so that appropriate repair mechanisms (e.g. reprompting) can be invoked if some form of error was likely to have occurred (e.g. a misrecognition, or unexpected utterance by the caller which was not modelled in the recognition grammar). Of course, the response time of the recognition module itself is also an important factor in the overall speed of the DE system (and even more to the caller's perception of speed).

Multilingual environments present additional challenges to systems: both the target names and the system users may be from very diverse language backgrounds. Pronunciation is directly influenced by both the language origin of the word and the mother tongue of the speaker [9], [10]. When words are mispronounced, speakers with the same language backgrounds tend to make similar errors when pronouncing names cross-lingually, an observation that has also been made for South African speakers [11]. Dealing with the resulting pronunciation variability is an area of ongoing research.

C. Language modeling

The language modeling task is critical to the success of automated DE. This is well illustrated in a US DE case study [5] where it is shown that phrase accuracy can range from unusable when using a flat unigram language model (under 4%) to highly accurate (94%) when using sophisticated, highly optimized language models.

From a user interface perspective, a more natural dialogue typically requires a more general, single language model, whereas the traditional directed-initiative user interface lends itself to a hierarchy of separate language models. A hierarchical language modeling approach is often tightly coupled with a directed-initiative DE dialogue, where the user is prompted for specific information. This approach is still considered to be the most popular in the US [8], for reasons which include the fact that callers are familiar with this approach and that local language models can be employed.

Several different types of local language models can be used. On one hand, language models can be estimated on data from a particular city or state as indicated in the listing database. The disadvantage of this approach is that the language models then vary significantly in size given the city and state. They can be too large for some cities such as Johannesburg and too small for towns such as Volksrust (using South African cities and towns as examples). An alternative approach may be to build language models based on local service areas (LSAs) if this information is available [8]. In this fashion Sandton and Alberton will have separate local language models, whereas Utrecht, Newcastle and Volksrust may share a local language model. Local language models are rarely used independently, as the literature consistently shows that interpolation with a larger national language model yields the best performing systems [4], [8], [12], [13].

Single language modelling approaches are popular when a less restrictive user-interface is employed. Other motivating factors for a single language model include simplicity (it is easier to optimize a single large language model than having to maintain and rebuild tens of thousands of language models) and scalability (a single language model can be loaded into memory and used by all incoming calls). Promising results have been obtained with a single class-based language model, which combines a city-state and listing language model with a filler plug-in model eg 'I'm looking for [[listing]] in [[city-state]]' [12].

III. Implementation

The demonstration application is a DE system supplying the primary telephone numbers of all South African municipalities to callers. A directed dialogue is used, and a single (combined) language model is used for all municipal names. Additional hand-crafted language models are used for a small set of disambiguations required. Confidence scoring is used to determine the need for reprompting and confirmation.

The multilingual DE application was implemented using open-source software where possible, and writing additional

open-source software where necessary. The telephony application was written in Asterisk Gateway Interface (AGI) and Perl, and Media Resource Control Protocol (MRCP) requests are possible via UniMRCP[1]. A Kaldi [14] UniMRCP plugin was written in order to make speech recognition with Kaldi possible with MRCP calls.

A. Decoder

An online Kaldi decoder, developed by Povey, was incorporated into the UniMRCP framework by means of a UniMRCP decoder plugin. The decoder generates lattices, which are used to obtain confidence scores for hypotheses by doing minimum Bayes risk (MBR) decoding.

Acoustic models were trained on a combination of corpora:

- Lwazi corpus [15]: the Afrikaans, English, Sesotho and isiZulu Lwazi corpora. This subset of the Lwazi corpus consists of 800 speakers, and amounts to 25 hours of speech.
- SADE corpus [16]: the entire SADE corpus, including an in-house subset that will not be released in the public domain, was used. The corpus contains 24 hours of speech from 44 speakers (a balance of Afrikaans, English, Sesotho and isiZulu male and female speakers).
- Municipality names corpus: this is an in-house corpus of spoken municipality names. The corpus consists of 2h30m of speech from 24 speakers.

The acoustic models were trained using a recipe similar to the Kaldi Babel & Wall Street Journal (WSJ) recipes; standard 3-state left to right triphone hidden Markov models (HMMs) (with Gaussian mixture models (GMMs) as the statistical model) were trained, with maximum likelihood linear transform (MLLT) and feature-space maximum likelihood linear regression (fMLLR) speaker-specific transforms. The features employed are standard Mel-frequency cepstral coefficients (MFCCs) with cepstral mean normalization (CMN) per speaker. Frames are spliced together, and linear discriminant analysis (LDA) is used to reduce the dimensionality of the features to 40. These models were then used to create alignments, which are used to initialize training of a deep neural network (DNN) with 3 hidden layers. The DNN is then used together with a hand crafted pronunciation dictionary and a flat municipality term grammar (273 municipality names) to create a decoding network which is used during online decoding.

B. User interface

The user interface employed in the current system is fully automated with no operator fallback. Callers are welcomed and then asked to say the name of the municipality they are looking for. The system recognises the speech and determines whether or not one of 273 municipalities were asked for. If the system is confident that it recognised the speech correctly (high confidence), the number for the municipality is provided to the caller. When the system is slightly unsure (medium confidence), it asks the caller to confirm whether or not it recognised the right municipality. If the system could not understand the caller (low confidence), it asks the caller to repeat the request. See Table I for examples of typical dialogues.

TABLE I. THE EFFECT OF CONFIDENCE SCORE ON CONFIRMATION DIALOGUE.

Example of a **high confidence** dialogue:	
SADE	Welcome to municipal directory enquiries! Please say the name of the municipality you are looking for.
Caller	City of Johannesburg.
SADE	The number for City of Johannesburg is 011 407 6111.
Example of a **medium confidence** dialogue:	
SADE	Welcome to municipal directory enquiries! Please say the name of the municipality you are looking for.
Caller	Matlosana
SADE	I think you said City of Matlosana. Is this correct?
Caller	Yes.
SADE	The number for City of Matlosana is 018 487 8300.
Example of a **low confidence** dialogue:	
SADE	Welcome to municipal directory enquiries! Please say the name of the municipality you are looking for.
Caller	Lesedi [background speech]
SADE	I did not understand what you said. Please say the name of the municipality you are looking for.
Caller	Lesedi.
SADE	The number for Lesedi is 016 340 4300.

IV. USABILITY TESTING

Informal evaluation of the application and its components occurred throughout the project, and culminated in a formal evaluation to determine the usability of the final DE system from a user perspective. These tests were conducted by allowing respondents to interact directly with the live telephony application. Written instructions were provided to respondents, who were then required to call a specific telephone number in order to execute a number of tasks. Once completed, respondents were asked to complete their response forms and return these to the analysts.

It is worth noting that the usability tests were conducted in two phases:

- Group 1: A short test phase (6 respondents: speakers 18-23) was conducted first. This was used to verify that the evaluation protocol itself produced sensible results, and to obtain a first set of formal responses from respondents. Based on the feedback obtained, two changes were made to the system. (See Section V-C.)
- Group 2: Once the system had been updated, a second and final phase (17 respondents: speakers 01-17) was conducted. The system used by these respondents is the one included in the final set of SADE deliverables.

The test protocol and respondent selection process are described in more detail below.

A. Usability evaluation protocol

The main aim of the usability testing was to determine whether the system is usable to a wide variety of users. This was determined in two ways: (1) by providing system users with very specific tasks to complete and measuring their ability

[1] http://www.unimrcp.org

to complete these tasks, and by (2) eliciting user feedback with regard to their interaction with the system.

The first aspect of the evaluation (functional task completion) also had two components: ten questions were 'directed tasks', asking the user to obtain a specific telephone number; five more were 'open tasks', asking the user to obtain the telephone number of any of a given set of municipalities. In both cases, users were asked to indicate whether the system provided the correct (expected) number. User feedback was then corroborated with information logged by the SADE system in order to prevent 'friendly users' from influencing results. The specific names offered for the open task were selected randomly from the 273 municipality names. Nine of the ten directed task names were also selected randomly: only the first name ('City of Johannesburg') was intentionally selected to be a name that would be familiar to most respondents.

The second aspect of the evaluation requested more subjective opinions from users. Users were asked to indicate their opinion with regard to the ease of use of the system, specifically by answering the questions listed in Table II below. An opportunity to comment on any aspect of the system was also provided.

TABLE II. QUESTIONS USED TO ELICIT SUBJECTIVE FEEDBACK DURING USABILITY TESTING.

On a scale of 1 to 5:
Did you find the system easy or difficult to use? *(1 = very difficult; 2 = difficult; 4 = easy; 5 = very easy)*
Did the automated system speak too fast, too slow, or was the speed right for you? *(1 = much too slow; 3 = perfect; 5 = much too fast)*
Did you find the system friendly or was it intimidating in any way? *(1 = intimidating; 3 = cannot really say; 5 = friendly)*
Do you prefer the telephone number to be spoken faster, slower or exactly as it is? *(1 = faster; 3 = no problem as it is; 5 = slower)*

B. Usability test participants

Twenty-three speakers were asked to participate in the final usability tests, with the majority being university students. Table III shows a language and gender breakdown of the participants.

TABLE III. LANGUAGE AND GENDER OF FINAL USABILITY TEST PARTICIPANTS.

Language	Male	Female	Total
Afrikaans	3	3	6
English	2	3	5
isiXhosa	0	1	1
isiZulu	3	3	6
Sesotho	2	3	5
Total	10	13	23

Of the twenty-three participants, six (two from Group 1 and four from Group 2) indicated that they had never used a telephone-based DE system before.

V. ANALYSIS OF RESULTS

A. Task completion rate

The primary question posed during usability testing is: 'Is a user able to obtain the number of a specific municipality, using only the SADE system?' We therefore analyse task completion rate for both the directed and undirected tasks. Results are provided in Table IV.

TABLE IV. AVERAGE TASK COMPLETION RATE.

	Directed tasks	**Open tasks**
Group 1	93.33	90.00
Group 2	95.88	98.82

In four cases, an incorrect number was played back to the user. As the dialogue (see Table I) explicitly mentions the name of the municipality for which the number is provided, incorrect number playback is automatically flagged to the users, allowing them to re-use the system if so required.

B. Recognition results

All 512 spoken utterances recorded during the usability tests were manually transcribed and compared with the online recognition hypotheses. In the following sections, a detailed analysis of the spoken utterances is presented.

1) Invalid utterances: Of the 512 recorded utterances, 60 were deemed to be invalid from a speech recognition perspective. (Note that such utterances may still be valid from a system usage perspective.) Reasons provided by the transcriber for invalidating utterances included:

- partially pronounced spoken municipalities, such as 'theewaters-',
- spoken noise (including significant background speech),
- empty utterances and utterances containing only noise,
- incomprehensible or 'out of vocabulary' utterances (eg 'zutangola'),
- three utterances where users rendered incorrect versions of the municipality name: 'joburg' and 'johannesburg', instead of 'city of johannesburg' and 'matlosana', instead of 'city of matlosana', and
- two utterances where the same user rendered unexpected variations of 'yes' and 'no' (using 'correct' instead of 'yes' and 'incorrect' instead of 'no').

Of the 60 invalid utterances:

- 46 were correctly rejected by the DE system,
- 6 were referred to the user for confirmation,
- 4 resulted in partial matches (with the call subsequently being ended, seemingly successfully),
- in 4 cases a number was incorrectly played back to the user, and
- in the case of 'incorrect' (for 'no') and 'correct' (for 'yes'), the correct number was played back, even though the system does not cater for those renditions of confirmations.

For the purposes of calculating recognition accuracy, the invalid utterances will be ignored as it is unclear what the user intent was, hence determining a ground truth is not always possible.

2) Total error rate: The total error rate was calculated on the 452 valid utterances from the final usability tests. Of these utterances, 418 were municipality name utterances (Table V), and 34 were yes/no utterances (Table VI).

TABLE V. TERM ERROR RATE FOR THE HIGH, MEDIUM AND LOW CONFIDENCE HYPOTHESES USING THE MUNICIPALITY RECOGNITION GRAMMAR.

Confidence	Number of utterances	Term error rate
High	327	3.67
Medium	43	23.26
Low	48	62.50

TABLE VI. TERM ERROR RATE FOR THE HIGH AND LOW CONFIDENCE HYPOTHESES USING THE YES/NO RECOGNITION GRAMMAR.

Confidence	Number of utterances	Term error rate
High	23	26.09
Low	11	54.54

3) Speaker error rate: Speaker error rates (calculated on all valid utterances) are shown in Figure 1 (municipality name recognition) and Figure 2 (yes/no recognition). No bar for a speaker indicates 100% accuracy (see for example speakers 01, 02, 06, 13, 14, 15, 16 and 20 in Figure 1). Also note that not all speakers had to confirm an utterance, hence the reduced number of speakers in Figure 2.

Figure 1 is interesting for several reasons. Firstly, it shows that almost all speakers achieved a high term recognition accuracy if the confidence was high. It is also evident that a couple of speakers struggled quite significantly with some terms. Speaker 11 stands out, and further analysis suggests that the speaker (a Zulu female speaker, who indicated that she has never used a DE system before), struggled significantly with the pronunciation of the two Afrikaans/Dutch municipalities for which she had to find the number (Theewaterskloof and Witzenberg). Speaker 17 (female Sotho speaker) also struggled with Afrikaans/Dutch municipality pronunciations (Theewaterskloof, Witzenberg and Overberg).

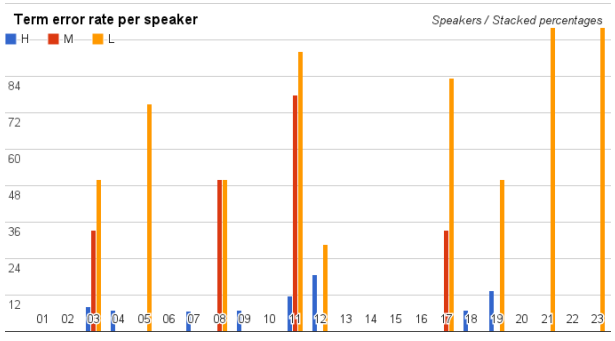

Fig. 1. Bar chart indicating the term error rate for each of the different confidence intervals per speaker for the municipality name recognition task. (Confidence indicated as *H*igh, *M*edium or *L*ow.)

C. User feedback

As mentioned in Section IV, the final usability tests reported on here were conducted in two batches: Group 1 was recorded first, and their feedback regarding the speed at which telephone numbers are read were considered important enough

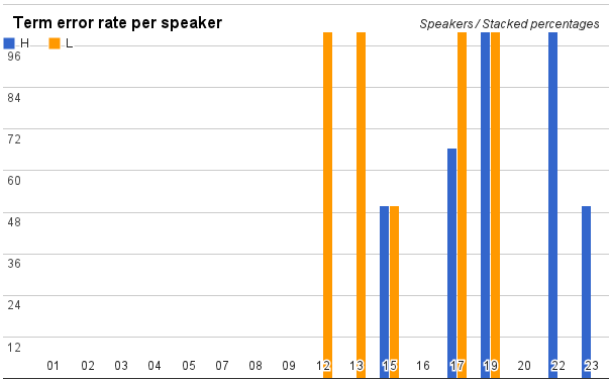

Fig. 2. Bar chart indicating the term error rate for each of the different confidence intervals per speaker for the yes/no recognition task. (Confidence indicated as *H*igh or *L*ow.)

that we decided to first incorporate this feedback before asking Group 2 to complete the test. Usability feedback will thus be reported separately for Group 1 and Group 2.

In addition to determining the usability of the system by measuring task success, several questions were posed to participants in order to gain further insight into general usability. The questions asked, as well as the feedback are displayed in Figures 3 to 5.

All the Figures confirm that Group 2 was slightly more positive about the system than Group 1, especially regarding the speed at which telephone numbers are spoken (Figure 5 does however also confirm that it is impossible to satisfy all user preferences).

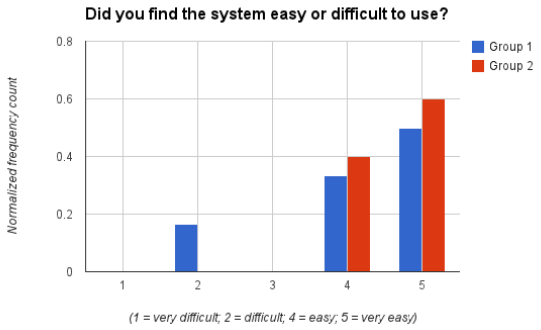

Fig. 3. Feedback per group on ease of use of the DE system.

D. Discussion

Overall results were very encouraging. All 23 speakers were able to perform multiple successful directory enquiries. User feedback showed that 95.65% of all tasks were completed successfully. Only 13% of calls were rejected, with a further 9.5% confirmed with the user. Even if all calls deemed invalid by the manual transcription process are included in the 'rejected' calls count, the percentage of 'rejected' calls is still only 23.24%. 96.33% of the accepted municipality names were correctly classified. This number drops slightly to 94.86% when taking 'yes/no' recognition results into account as well.

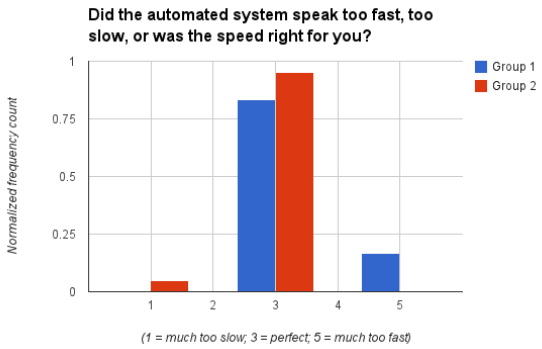

Fig. 4. Feedback per group on speed of the DE system persona's speech.

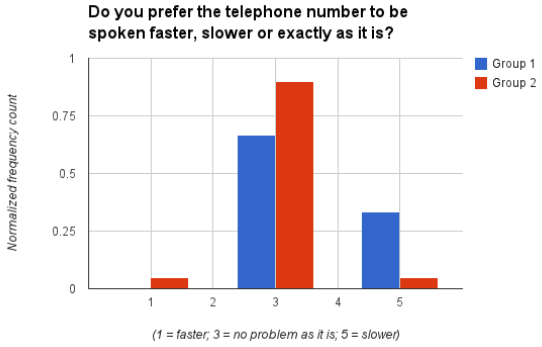

Fig. 5. Feedback per group on speed at which telephone numbers are spoken.

While we deem the current system ready for general use, future work will include:

- extending the grammar to include additional alternative spoken forms for a single concept,
- improving the yes/no recognition,
- implementing session-based CMN instead of utterance-based CMN,
- performing speaker adaptive training; this will however have a significant influence on the required disk usage,
- evaluating the system with speakers of all of the eleven official South African languages and subsequent analysis of recognition results, and
- a comprehensive language-specific analysis of recognition results.

VI. Conclusion

Practical DE systems that serve names from multiple languages to speakers of multiple languages are still rare. We review the development of DE systems in general, and demonstrate how a 'Municipality Hotline' developed according to currently available technologies can provide a practically usable service, even when both the task and the speakers represent diverse language backgrounds. Formally evaluated with first language speakers from four distinct South African languages (Afrikaans, English, isiZulu and Sesotho), an overall task completion rate of 95.65% was achieved, with the large majority of users finding the system simple and easy to use. While the dialogue used here was still fairly restricted, these results bode well for the development of more extensive DE systems.

Acknowledgement

This research was co-sponsored by the Department of Arts and Culture of the Republic of South Africa, Molo Innovations, IntSyst Labs and the Technology Innovation Agency. We would also like to thank the rest of the SADE application team (Derik Thirion, Anina Lambrechts, Aditi Sharma and Willem Basson) for their various contributions to the development and evaluation of this system.

References

[1] H. M. Chang, "Comparing machine and human performance for callers directory assistance requests." *International Journal of Speech Technology*, vol. 10, no. 2–3, pp. 75–87, 2007.

[2] C. A. Kamm, C. R. Shamieh, and S. Singhal, "Speech recognition issues for directory assistance applications." *Speech Communication*, vol. 17, no. 3, pp. 303–311, 1995.

[3] E. Barnard, J. Schalkwyk, C. van Heerden, and P. J. Moreno, "Voice search for development," in *Proc. Interspeech*, Makuhari, Japan, 2010, pp. 282 – 285.

[4] D. Yu, Y. C. Ju, Y. Y. Wang, G. Zweig, and A. Acero, "Automated directory assistance system – from theory to practice," in *Proc. Interspeech*, Antwerp, Belgium, 2007, pp. 2709–2711.

[5] A. Moreno-Daniel, J. Wilpon, and B. H. Juang, "Index-based incremental language model for scalable directory assistance," *Speech Communication*, vol. 54, no. 3, pp. 351–367, 2012.

[6] M. Lennig, G. Bielby, and J. Massicotte, "Directory assistance automation in Bell Canada: Trial results," *Speech Communication*, vol. 17, no. 3, pp. 227–234, 1995.

[7] G. Lindgaard and D. Caple, "A case study in iterative keyboard design using participatory design techniques," *Applied ergonomics*, vol. 32, no. 1, pp. 71–80, 2001.

[8] E. Levin and A. M. Man, "Voice user interface design for automated directory assistance," in *Proc. Interspeech*, Lisboa, Portugal, 2005, pp. 2509–2512.

[9] A. F. Llitjós and A. W. Black, "Knowledge of Language Origin Improves Pronunciation Accuracy of Proper Names," in *Eurospeech*, 2001, pp. 1919–1922.

[10] B. Réveil, J.-P. Martens, and B. D'Hoore, "How speaker tongue and name source language affect the automatic recognition of spoken names," in *Proc. Interspeech*, Brighton, UK, 2009, pp. 2971–2974.

[11] M. Kgampe and M. H. Davel, "The predictability of name pronunciation errors in four South African languages," in *Proc. PRASA*, Vanderbijlpark, South Africa, Nov. 2011, pp. 85–90.

[12] C. Van Heerden, J. Schalkwyk, and B. Strope, "Language modeling for what-with-where on goog-411," in *Proc. Interspeech*, Brighton, UK, 2009, pp. 991–994.

[13] A. Stent, I. Zeljkovi, D. Caseiro, and J. Wilpon, "Geo-centric language models for local business voice search," in *Proc. NAACL*, May 2009, pp. 389–396.

[14] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The Kaldi Speech Recognition Toolkit," in *Proc. IEEE 2011 Workshop on Automatic Speech Recognition and Understanding (ASRU)*, Dec. 2011.

[15] E. Barnard, M. Davel, and C. van Heerden, "ASR corpus design for resource-scarce languages," in *Proc. Interspeech*, Brighton, UK, Sept. 2009, pp. 2847–2850.

[16] J. W. Thirion, C. van Heerden, O. Giwa, and M. H. Davel, "The South African Directory Enquiries (SADE) corpus," to be submitted.