Voice Interaction Design

The Journal of AVIxD Association for Voice Interaction Design

Vol. 1, Issue 2, November 2016 pp. 1-8

Designing and Developing Solutions for Automated Name Capture

Dmitry Sityaev
Senior Speech Scientist
Genesys
2001 Junipero Serra Blvd
#600
Daly City, CA
USA
dmitry.sityaev@genesys.com

Abstract

Recognition of proper names presents many challenges for speech recognition. This paper addresses the issues encountered when building name capture solutions. Various recognition techniques are examined that help to improve recognition accuracy and coverage. A sample dialogue is provided that uses a novel fallback strategy aimed at increasing the task completion rate for name capture.

Keywords

Speech recognition, name recognition, grammars, corpora

Introduction

Name capture is one of the most difficult tasks that developers and designers have had to address when developing interactive voice response (IVR) solutions. From dealing with constraints imposed by recognition technology to devising solutions that provide an enhanced customer experience, name recognition presents many challenges that must be addressed to build and deploy successful applications. This paper throws some light on the challenges presented by name recognition and offers some guidance for developers and designers on how to approach the creation of such solutions.

First, we survey some of the literature that has investigated the recognition of proper names and addressed the usability side of name recognition. Then, we provide some direction as to where to look for name corpora, which is indispensable in building name capture solutions. Further, we explore different grammar writing techniques and strategies to report on their suitability for name capture. Finally, we present a proposed solution aimed at capturing names, followed by recommendations and conclusions.

Recognition of Proper Names

Recognition of proper names is a challenging task. There are various issues surrounding it. First, systems built to capture names are aimed at recognizing large sets of names, easily ranging into the tens of thousands.

Second, to achieve satisfactory recognition results, a high-quality pronunciation dictionary is required (especially for those languages where the grapheme-phoneme conversion is not one-to-one, e.g., English). These pronunciation dictionaries are often handcrafted and may not be readily available for all languages.

To complicate the matter further, names in some languages (e.g., English) are homophonous. For example, in some dialects of American English ['kɛri] can correspond to any of the following spellings: Carrie, Carie, Cary, Cari, Carey, Carry, Kerry, Karry, Kerrie, Kerri and so on. The system must be able to disambiguate such cases, a task that is not trivial.

One way to combat some of the aforementioned issues is to ask a person to spell their name; indeed, this is a strategy humans often resort to when they are unsure of the name spelling. Recognition of spelled names has been explored in many studies (Kamm, Shamieh, & Singhal, 1995; Meyer & Hild, 1997; Neubert, Gravier, Yvon, & Chollet, 1998; Seide & Kellner, 1997), and specifically, the recognition accuracy of spelled names has been shown to be more accurate than the recognition accuracy of spoken names. For example, Meyer and Hild (1997) report a recognition accuracy of 60.0% for spoken names vs. a recognition accuracy of 96.5% for spelled names on a test set of 1,337 German names.

Meyer and Hild (1997) also looked at combining the information from recognizing the spoken and spelled name into one utterance. The findings were that there was a small improvement in accuracy: the spelling remained the main source of information, whilst the information obtained from the spoken name yielded a slight improvement over the accuracy obtained from spelling alone.

Whereas the recognition of names has received a lot of attention in the literature, only a few studies have focused on the usability evaluation of solutions aimed at name capture. Davidson, McInnes and Jack (2004) explored three different strategies for capturing surnames: speak only, a one stage speak and spell (where callers spoke and spelt the surname in one utterance), and a two stage speak and spell (where callers spoke the name in a first utterance and then spelt their surname in a second one). There was a clear user preference for the Speak and Spell strategy over the Speak Only strategy. Although recognition accuracy figures for subjects providing their own surname were similar to those reported in Meyer and Hild study, overall task completion rates were lower. This is not surprising as task completion rates take into account both valid (in-grammar) and invalid (out-of-grammar) responses. Davidson, McInnes and Jack found that the Speak and Spell strategy yielded significantly higher task completion rates (there was no substantial difference between the one stage vs. two stage strategies) (Table 1).

Table 1. Summary of findings from Davidson, McInnes and Jack (2004)

Strategy	Mean attitude score	Explicit preference Most preferred	Least preferred	Task completion
Speak Only	4.57	13.7%	63.2%	51.6%
Speak and Spell (One Stage)	5.18	46.3%	10.5%	80.0%
Speak and Spell (Two Stages)	5.17	37.9%	17.9%	77.9%

In another paper, Damper and Gladstone (2007) remarked that in their usability assessment the users found it “extraordinarily hard, and unnatural, to speak one’s name followed by its spelling (e.g., 'John j o h n Smith s m i t h')”. So perhaps, from a usability perspective, it makes sense to collect first and last names in two separate dialogue turns.

Name Corpora

The availability of data plays a crucial role in language and speech research. Just as it is important to have acoustic data for training acoustic models for speech recognition, it is vital to have name corpora and name pronunciation dictionaries to build recognition solutions. This section suggests potential sources of name data, and highlights issues surrounding these sources.

One source of publically available information is records assembled and provided by government organizations. These may include marriage, birth, divorce, death records, etc. Of particular interest and use is census information. In most countries censuses are carried out on a regular basis (for example, the US Census is conducted every 10 years on the decade). The information disseminated may already be text pre-processed and may contain frequency counts and distributions (US Census, 1990; US Census, 2000). However, the information released to the public may not always be 100% complete (e.g., cutoff at 95% frequency counts), as there are often stringent requirements that data shall never identify any specific individual by name, address, etc. In some countries, the census data is kept away from the public for as long as 100 years (e.g., the UK).

Another source of name data is information available on the Internet. Name data can be extracted from personal profiles available through various professional and social networks. Such data may not always be ‘clean’ (e.g., may contain spam, orthographic errors, etc.) and may require several stages of cleaning after its collection. Moreover, it is also unclear as to how representative the data from such public profiles are of the whole population of a country.

Phonebooks have been used in the past to conduct research on names and their distribution (Rogers, 1995). They are known to be sources of good population coverage (e.g., Rogers reports coverage up to 80%). Phonebooks, however, are not without their faults; phonebooks covering different areas are not always issued simultaneously, they do not take into account modern-day practices of co-habitation, and individuals may choose to have their name and/or number not listed in the phonebook. But perhaps, most importantly, phonebooks are increasingly becoming obsolete in a digital world.

There are also providers of data resources who sell data either for commercial or research purposes. There is often a considerable cost involved here that needs to be taken into account. Some name corpora are available free-of-charge (e.g., Carvalho, Kiran, & Borthwick, 2012), however, these corpora are often restricted for research purposes only.

Therefore, it appears very difficult (if not impossible) to find name data sources that are 100% complete and accurate. As pointed out above, the distribution and frequency of names tends to change over time, which introduces the challenge of finding the most up-to-date data. Finally, availability of such data is often subject to various data protection acts that may limit the amount of data made available to the public.

Building Context-Free Grammars for Name Recognition

In this section, we will explore how different grammar building techniques and approaches affect recognition performance for names. These techniques are not exclusive for the recognition of names, and can be applied in other recognition tasks.

For the purposes of experimentation, the data from 1990 US Census were used for building context-free grammars. This corpus consisted of 1219 unique male first names, 4275 unique female first names and 88799 unique last names.

The test data were obtained from a live application and comprised 1268 spelled first names and 1261 spelled last names.

Grammar performance is reported using the following two (routinely used) metrics:

- coverage: how many utterances are covered by a grammar in question across all utterances observed
- accuracy: how many utterances are correctly recognized by a recognizer across all utterances observed

The above accuracy metric is often referred to as 'perceived grammar accuracy'. There is another metric commonly known as 'in-grammar accuracy' where correctly recognized utterances are measured as a percentage of in-grammar utterances only. Perceived grammar accuracy is reported in the experiments discussed below.

Applying constraint

In Experiment 1, grammars were constructed covering any sequence of letters between 2 and 12 letters. As can be seen from Table 2 and Table 3 below, although the coverage is high, the accuracy is very low for such unconstrained grammars.

The picture changes drastically when a grammar is built from the corpus data. The results from Experiment 2 show that accuracy goes up significantly when the grammars are built using the corpus data; there is, however, a decrease in coverage.

To see if the grammar coverage can be maintained, an experiment was run where the two grammars from Experiment 1 and Experiment 2 (for first names and last names, respectively) were run in parallel (Experiment 3). Despite the grammar coverage being at the same level as in Experiment 1, the accuracy dropped significantly compared to the accuracy obtained in Experiment 2.

Weighting

Further experimentation was carried out to see whether grammar weighting would improve recognition accuracy. A series of experiments were performed (see Experiment 4 below) in which the two grammars from Experiment 1 and Experiment 2 were run in parallel, with different weights assigned to both grammars. As the grammar producing an unconstrained sequence of letters was de-weighted, the recognition accuracy steadily increased.

In Experiment 5 the items (e.g., spelled names) within the grammar created from the Census corpus were weighted based on their frequency (as provided in the 1990 Census corpus). For first names, weighting contributed to an improvement, yielding the best accuracy across all runs. Weighting did not improve the accuracy of last names.

Grammar size

Another dimension that can be explored for improving accuracy is varying the grammar size. Grammar accuracy on the same test set will decrease as the grammar size is increased. Several grammars for last names were created in Experiment 6 out of the corpus data (for top most frequent names) which varied in the number of names. Table 2 shows the experimental results. Taking into the account the trade-off between coverage and accuracy, the optimal grammar size was 25000 names, yielding highest accuracy across all runs and providing a tolerable decrease in coverage. No corresponding experiments were performed for first names as the size of the first name corpus was relatively small.

Optimization and other techniques
Further grammar accuracy optimizations are possible, including, but not limited to, pronunciation modelling and experimentation with various recognizer parameters. Results are not reported here as gains may vary from one recognizer to another, and will depend on the arsenal of recognition features available for control.

Table 2. Coverage and accuracy results from the experiments on first names

Experiment	Description	Coverage	Accuracy
Experiment 1	letter grammar	86.0%	18.9%
Experiment 2	corpus grammar	56.1%	44.7%
Experiment 3	corpus + letter grammar	86.0%	19.3%
Experiment 4	corpus + letter grammar weighted		
	corpus: 0.95, letter: 0.05	86.0%	31.2%
	corpus: 0.995, letter: 0.005	86.0%	36.7%
	corpus: 0.999, letter: 0.001	86.0%	40.2%
	corpus: 0.9999, letter: 0.0001	86.0%	43.2%
	corpus: 0.99999, letter: 0.00001	86.0%	44.7%
Experiment 5	corpus grammar weighted	56.1%	47.3%
Experiment 6	corpus grammar size	not run	not run

Table 3. Coverage and accuracy results from the experiments on last names

Experiment	Description	Coverage	Accuracy
Experiment 1	letter grammar	93.5%	18.7%
Experiment 2	corpus grammar	73.1%	50.5%
Experiment 3	corpus + letter grammar	93.5%	19.4%
Experiment 4	corpus + letter grammar weighted		
	corpus: 0.95, letter: 0.05	93.5%	33.2%
	corpus: 0.995, letter: 0.005	93.5%	42.4%
	corpus: 0.999, letter: 0.001	93.5%	47.3%
	corpus: 0.9999, letter: 0.0001	93.5%	52.3%
	corpus: 0.99999, letter: 0.00001	93.5%	52.7%
Experiment 5	corpus grammar weighted	73.1%	34.4%
Experiment 6	corpus grammar size		
	70000 names	71.8%	51.3%
	50000 names	70.8%	52.6%
	25000 names	68.2%	54.5%
	10000 names	62.6%	53.0%
	5000 names	57.3%	50.7%

Proposed Design and Pilot Evaluation

Previous sections have reported on the results of proper name recognition, usability studies involving the solutions for name capture, and strategies for building grammars. In this section, we will revisit some of the design decisions and propose a solution for a novel name capture process.

As shown above, the grammars created never reached 100% coverage. Not only is there a trade-off between coverage and accuracy (as we vary the grammar size for example), there is also the coverage limitation in the corpus data due to privacy concerns. Additionally, no recognition is 100% accurate, and one must also address the issue of false accepts and false rejects.

It is possible to use a spelling strategy to capture names not covered by the grammar. Using spelling at the fallback stage has been explored in a few studies (Bauer & Junkawitsch, 1999; Córdoba, San-Segundo, Montero, Colás, Ferreiros, Macías-Guarasa, & Pardo, 2001; San-Segundo, Colás, Córdoba, & Pardo, 2002); it has been shown that adding a fallback step with the spelling strategy can substantially increase the total number of names captured successfully. Additionally, it was decided to split the name collection at fallback into chunks: recognizing chunks of 3-4 letters can lead to better recognition results than recognizing long sequences of letters.

A sample dialogue of what the callflow may look like is provided below. Other dialogue designs are possible, of course, which may incorporate more sophisticated solutions (e.g., n-best); however, these are not explored here and point to an area of future research.

IVR: Please spell your first name
User: J, O, N, A, T, H, A, N
IVR: I heard Jonathan. J, O, N, A, T, H, A, N. Is that right?
User: Yes
IVR Great. Now, please spell your last name
User: H, A, V, E, N, H, I, L, L
IVR: I heard Havener. H, A, V, E, N, E, R. Is that right?
User: No
IVR: Sorry about that. Let's try this a different way. How many letters are in your last name?
User: Nine
IVR: Ok, nine. Let's enter them in groups. What are the first three letters?
User: H, A, V
IVR: I heard H, A, V. Is that correct?
User: Yes.
IVR: And what the next three letters?
User: E, N, H
IVR: I heard E, N, H. Is that right?
User: Yes.
IVR: Finally, what are the last three letters?
User: I, L, L
IVR: I heard I, L, L. Is that correct?
User: Yes
IVR: Thanks

A small pilot study was conducted in which 100 callers provided their first name and last name. The findings revealed that using this fallback strategy when there were misrecognitions or coverage issues increased the task completion rate by 20-25%.

Recommendations

Below are recommendations for speech scientists and VUI designers that provide some guidance when building solutions involving name capture:

- Invest time and money into researching and obtaining name corpora with good coverage and name dictionaries with accurate pronunciations.
- Experiment with different grammar techniques and data sources to achieve the best recognition accuracy and coverage.

- Use spelling as a powerful strategy when capturing names.
- Consider designing and implementing a fallback approach to handle names not covered by the grammar or in case the recognizer is struggling.
- Carry out usability studies to investigate callers' reaction to the chosen strategy for name capture.

Conclusion

This paper has provided some insights into what it takes to build a solution for name recognition. It was shown that spelling is a powerful tool that can be used when capturing someone's name. Various grammar writing techniques were explored and it was shown how they can be used to increase recognition accuracy. The importance of carrying out usability studies and researching name corpora was also highlighted.

Future directions for this research will focus on exploring other methods for constraining recognition (e.g., statistical language modelling) as well as experimenting with improved acoustic models for alphanumeric capture. It would be interesting to extend the proposed callflow to the Speak and Spell strategy and compare performance results to those obtained in Davidson, McInnes and Jack (2004). Also, the proposed sample dialogue calls for a usability study, focusing specifically on the fallback approach and the use of the chunking strategy.

References

Bauer, J., & Junkawitsch, J. (1999). Accurate recognition of city names with spelling as a fallback strategy. *Proceedings of Eurospeech* (pp. 263-266). Budapest, Hungary: ISCA.

Carvalho, V.R., Kiran, Y., & Borthwick, A. (2012). The Intelius nickname collection: quantitative analysis from billions of public records. *Proceedings of NACL HLT*. Montreal, Canada: Association for Computational Linguistics.

Córdoba, R., San-Segundo, R., Montero, J.M., Colás, J., Ferreiros, J., Macías-Guarasa, J., & Pardo, J.M. (2001). An interactive directory assistance service for Spanish with large-vocabulary recognition. *Proceedings Eurospeech* (pp. 1279-1282). Aalborg, Denmark: ISCA.

Damper, R.I., & Gladstone, K. (2007). Experiences of usability evaluation of the IMAGINE speech-based interaction system. *International Journal of Speech Technology, 9,* 41-50.

Davidson, N., McInnes, F., & Jack, M.A. (2004). Usability of dialogue design strategies for automated surname capture. *Speech Communication, 43,* 55-70.

Kamm, C.A., Shamieh, C.R., & Singhal, S. (1995). Speech Recognition issues for directory assistance application. *Speech Communication, 17,* 303-311.

Meyer, M., & Hild, H. (1997). Recognition of spoken and spelled proper name. *Proceedings of Eurospeech* (pp. 1579-1582). Rhodes, Greece: ISCA.

Neubert, F., Gravier, G., Yvon, F., & Chollet, G. (1998). Directory name retrieval over the telephone in the Picasso project. *Proceedings of the IEEE IVTTA Workshop. Torino, Italy: IEEE.*

Rogers, C. (1995). *The surname detective: investigating surname distribution in England, 1086-present day*. Manchester: Manchester University Press.

San-Segundo, R., Colás, J., Córdoba, R., & Pardo, J.M. (2002). Spanish recognizer of continuously spelled names over the telephone. *Speech Communication, 38,* 287-303.

Seide, F., & Kellner, R. (1997). Towards an automated directory information system. *Proceedings of Eurospeech* (pp. 1327-1330). Rhodes, Greece: ISCA.

US Census 1990. Frequently occurring surnames from the Census 1990. Available at http://www.census.gov/topics/population/genealogy/data/1990_census.html.

US Census 2000. Frequently occurring surnames from the Census 2000. Available at http://www.census.gov/topics/population/genealogy/data/2000_surnames.html.

About the Author

Dmitry Sityaev
Dmitry is a Senior Speech Scientist at Genesys. He has presented and published papers in the areas of speech technology evaluation and prosody research.