

NAME DIALING – HOW USEFUL IS IT?

Kari Laurila, Petri Haavisto

Nokia Research Center, Speech and Audio Systems Laboratory, Tampere, Finland
Email: {kari.laurila, petri.haavisto}@nokia.com

ABSTRACT

Progress in automatic speech recognition technology has resulted in an increasing amount of deployed applications. Typically, the measure of success has been the amount of deployed or sold units and it has been much more difficult to evaluate the real user benefit from the technology. Actually, the usability, or usefulness, has largely remained an open issue. In this paper we focus on name dialing and discuss its usefulness from different angles, with a strong emphasis on mass market use and inexperienced users. As a concept, name dialing brings us back from where telephony started: an operator assisted way of making calls without a need to remember numbers. In essence, name dialing offers a solution to a minor inconvenience – using the directory. Even the arguably biggest advantage of name dialing, simplified car usage, is still less significant than the various concerns of users. Until time and further technical improvements alleviate the main concerns, usage of name dialing will remain as an occasional, rather than a primary, way of making calls.

1. INTRODUCTION

To most people, speech recognition is still a brand new technology that they rarely come across in every-day life, and, hence, one of the first worries in starting to use it is the recognition accuracy. The first impression often determines whether or not one is going to start using a system. It is rather easy to quote 95% recognition accuracy as being the acceptance threshold, but one still has to understand that different applications require different performance levels in order to be accepted and taken into use.

To give an illustrating acceptance example from the recognition accuracy point of view, we tried out a commercial dictation system. After the speaker adaptation session, we dictated some test sentences to see how the system works. Below, the correct dictated sentences are given first, followed by the recognition result (*in italic*).

Dear IEEE Member: Your membership in the IEEE and your professional profile tell me you are an exceptional candidate for membership in the IEEE Computer Society. You are a serious professional concerned about your career and lifelong learning. You understand the importance of technical competence and receiving up-to-the-minute information about developments in computing – the type of information enjoyed by Computer Society members.

Your eye Tripoli member: You meant this is being done at Tripoli and your profession of profile tell me you are an exception of a candidate for membership in the right replete Computer Society. Your serious professional concerned about your career and lifelong learning. You understand the importance of technical competence and receiving up-to-the-minute information about developments in

building – the type of information in charge by Computer Society members.

From the above, one can see that the dictation system recognized many words correctly, but on the other hand, some words incorrectly. However, each reader may still have different opinions about its usefulness. Some may say that the system seems to make too many errors. Some may think that for them the system might work adequately to provide an initial draft for manual editing. We acknowledge that there are many factors that explain the mediocre performance of the recognizer in this situation: the authors are not native English speakers, the topic area is probably not typical to what the recognizer works best with, and the text includes many words which most likely do not appear in the dictionary of the system. This example takes us to consider the ease of error correction and its implications on usefulness, which already leads us away from the actual recognition accuracy as being the only merit of usefulness.

Since the determination of whether something is useful or not is generally very complex, we are not trying to answer whether speech recognition as such is useful or not (naturally we think it is very useful), but we focus on name dialing alone. Name dialing is often referred to as an important application offering a fast and easy way to set up calls, with no need to remember (or type in) telephone numbers. Having said this, one must admit that human factors like user acceptance, ease-of-use, and learnability still need further evaluation for various name dialing deployments.

In this paper, we try to approach name dialing from various angles and give some insights for the readers. Our main viewpoint is that of mass-market consumer applications and our arguments assume that the systems will be used essentially by everyone. First, we remind readers that with name dialing we are actually going back to the past using future means. After this, we continue by introducing speaker-dependent and speaker-independent name dialing concepts followed by some experiments that we have carried out. Then we discuss advantages and concerns of name dialing, as perceived by users. Before conclusions we give some thoughts for the way to make name dialing more useful in the future.

2. CLOSING CIRCLE

In early days of telephony, calling was accomplished with human operator assistance. One needed to call an operator and say to whom he or she wanted to talk to. Later on, together with technical progress, human operators were replaced with automatic telephone switches. Telephones started to have rotary discs which enabled callers to select numbers. Each person had a telephone number which consisted of a series of digits. When

a user selected a correct digit sequence the switch connected the caller to the desired person.

This telephone number oriented approach has ever since been the default way to make phone calls. However, users have always been calling to real persons having names, but instead of giving those names as input users have needed to input attached number sequences which they most often retrieve by browsing their personal phonebooks (typically in a paper format). In addition to automatic connections, human operator assisted calling has remained available, but often with some extra cost. The cost has, at least in some countries, been so high that perhaps only business users have resorted to this service on a regular basis.

The digital age together with mobile phones and other mobile devices brought electronic phonebooks from which names could be selected. Suddenly, the users were able to store names and corresponding telephone numbers and afterwards they were able to make calls by selecting the desired name from the phonebook. This has led us to a situation in which users remember ever fewer phone numbers. Today, a popular way to call is to find the desired name from the electronic phonebook and call the person. That is, the telephone number is stored once and it is invisible in the background for the rest of the lifetime of the system. Numbers are needed less frequently.

Recently, due to the rapid progress in semiconductor technology it has now become possible to integrate automatic speech recognition algorithms in consumer devices, e.g., mobile phones. Speech recognition is bringing us back to the starting configuration. Again, saying the name of the person the user would like to call is enough. The circle has closed. Due to technical progress, every user can have his or her own personal operator placed in the device itself, all the time waiting for the user to say names and make calls.

3. SPEAKER-DEPENDENT NAME DIALING

In speaker-dependent name dialing each user creates his or her own personal directory by training name tags attached to the corresponding telephone numbers. In training, the user is requested to speak each name one or more times. The speech recognition models are then constructed from these samples. The maximum amount of name tags typically varies from below ten to some tens of names. Users are free to train given names, full names (given name and surname), or something else. System manuals usually advice users to speak longer names having potentially higher recognition accuracy.

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 recognized name.

Speaker-dependent name dialing is widely available to the users, implemented either in a communication device itself or as a network service. For example, many mobile phone manufacturers are providing a speaker-dependent name dialing feature supporting ten name tags or more.

4. SPEAKER-INDEPENDENT NAME DIALING

The biggest weakness of speaker-dependent name dialing is quite clear: the required training phase. The training phase is an additional burden to the user, and the training is also prone to errors. In addition, speaker-dependency itself can be seen as a weakness. Speaker-dependent name dialing service is essentially for one person only.

Speaker-independent name dialing provides remedies for these problems. Basically, users can start making calls without a need to train the names first. In addition, the same name lists (and speech recognition models) can be used by multiple users.

Speaker-independent name dialing has already been implemented in some mobile phones for the Japanese market. The authors are not aware of other consumer device implementations. As a network offering, speaker-independent speech recognition technology enables a so-called corporate dialer service where employees of a company can call each others by simply speaking their names. The name list is the same for each user and it can be centrally maintained.

5. EXPERIMENT

During the last several years, we have carried out different experiments and test periods for different name dialing algorithms and systems. We have conducted novice and expert user tests, usability and recognition accuracy tests, terminal vs. network based name dialing tests, and so on. In this paper, we would like to describe one test [1] in more detail and finally present more general observations in the next section.

Test systems. Two different network-based name dialing systems were offered to the users. The first system was speaker-dependent, and the second system was speaker-independent. In the speaker-dependent system each user constructed his or her name list. In the speaker-independent system the basic name list offered to each user consisted of a selected group of Nokia employees, and in addition, each user was free to add more names via a web interface. Both systems had over 50 users. The systems could be accessed from any phone: office, home or mobile including hands-free car installations.

Database results. In order to get an idea about the basic recognition rates of the system engines, an off-line (non-telephony) 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 this experiment the recognition accuracy was measured with the original quiet environment samples and with noise corrupted samples (car noise 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.

	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 an off-line database.

Field test results. The average success rate obtained with the speaker-dependent system was 70% while for the speaker-independent system it was 84%. The success rate for the speaker-independent system was higher despite the fact that the base name list contained over 100 names. The results contradicted somewhat the database results which indicated that the speaker-dependent models perform better. However, we explain this result by the fact that speaker-dependent system users preferred to train names that were rather short (e.g. "Masa" instead of "Matti Nykänen"). In fact, 93% of the trained names consisted of a single word. Also, the training algorithm had accepted training utterances that contained extra noises and some utterances that were incorrectly end-pointed, resulting in truncated or extended name tags. Moreover, users were speaking more out-of-vocabulary names, partially due to the fact that they did not remember what names they had trained and how they had pronounced the names. The pre-defined name list in the speaker-independent system ensured the consistency of the models used for recognition.

Subjective opinions. Most of the users perceived the recognition accuracy as being very high, 81% and 83% of the users, respectively, gave the highest grade in their evaluation. Overall, both systems were perceived as useful (speaker-independent being more useful), easy-to-use, fast in dialing mode, and safe to use in a car. The biggest problem of the speaker-dependent system was the required training phase which was perceived slow. In the speaker-dependent case, half of the users expressed that 15 names or so was an adequate amount, while the other half did not accept any limits for the overall amount of names. Rejection of out-of-vocabulary names was perceived useful, especially among novice users. However, expert users were less in favor of sensitive rejection since the feature sometimes resulted in rejections of correct names.

6. NAME DIALING: ADVANTAGES AND CONCERNS

"Speech is the most natural way to exchange information between humans" is a widely used opening sentence in speech recognition related research articles, project plans, or proposals. However, to many, the use of speech recognition in every-day life remains dubious. The general public also remains ignorant of the performance and limitations of current ASR systems, which causes most of the recognition errors.

With mobile phones, name dialing offers a solution to a minor inconvenience – using the directory. Instead of pressing buttons to find the correct number from the phone memory, one can make a call by simply speaking the name of the called party. In essence, the advantage is small. Name dialing is capable of providing some saving from the dialing time, and in addition, it

increases the ease of dialing to a certain extent. In other words, users would be able to 'go through' the directory or call register more quickly and with less hassle than before. Considering this relatively small advantage, it is easy to accept that the usage of name dialing is currently not widespread, and the usage is only an occasional, rather than a primary, way of making calls.

One significant benefit of name dialing is perceived in car use by drivers. The benefit of not having to take your hands off the wheel and eyes off the road to make a call is significant. Thus, arguably the strongest appeal for the name dialing concept comes from those spending substantial amounts of phone time in the car.

In spite of the fact that there are some definite drivers for increased use of name dialing, there are also many concerns, which means that the overall response to name dialing has been more lukewarm than highly positive.

First of all, some level of embarrassment is evident. Embarrassment varies from high to low. Some users do not like to use the feature at all in front of other people, some use it in front of friends, and some use it in front of unfamiliar people without any problems. The embarrassment comes from the apparently ridiculous act of talking 'to' an inanimate object. This effect is clear even though people seem to have much less concern for talking to the phone in public once the call is established.

Using name dialing in a crowd draws attention to the user in a worse way than current phone usage does and one feels uncomfortable 'being stared at'. The current trend in earphone usage in public places is still strange to most, and name dialing is even more so. Having said this, cell phone usage was 'strange' at first but now it is a common site. As name dialing becomes a common feature, much of this embarrassment barrier is likely to go away.

Users also often appreciate the ability to make a call in private without anyone knowing who they are calling, even if they are surrounded by complete strangers. If you are 'announcing' to the world who you are calling (or worse still, dialing the number verbally) then there is a real concern that people will listen more intensively to what you are saying.

Many mobile phones support a so called 'speed dialing' feature. That is, with a long-press of a certain key, a call is made to the corresponding phone number. There are typically about 10 keys that are attached to speed dial numbers. For experienced speed dial users there appears to be little benefit with name dialing, since you still have to use your hands to activate it. In addition, users have justified doubts related to reliability: "Would it recognise your voice every time, without fail?", "Will it dial the right person?", "What if my voice changes due to a flu?", "Will it work in 'normal' (i.e. noisy) environments?" and so on.

Usability of name dialing significantly depends also on the environment, or the usage situation. Some places where name dialing is clearly acceptable include sitting in a car alone, skiing, in an office – privately, on a quiet street, alone at home. On the other hand, situations that may not be favourable for name dialing include traveling in a train, in a taxi, in office with

colleagues present, in a meeting, in a restaurant, and in an airport lounge.

7. TOWARDS BETTER NAME DIALING

Many of the current concerns of the users are such that only time and experience with speech recognition systems can alleviate them. However, there are some observations suggesting that shorter term improvements increasing usefulness are possible. In this section we would like to concentrate on one observation: the possibilities and challenges of the speaker-independent approach.

For a wider and more successful use of name dialing in the future, a ready-to-use speaker-independent version should be offered without any additional training phase. There are some feasible ways to achieve this. We will first describe an approach relying on implementation within a mobile terminal itself. Finally we discuss the network-based approach.

Terminal based solution. In mobile terminal, speaker-independent name dialing can be realized based on text fields (names) of memory entries. One can construct a speech recognition model if the written version of a name is available. The spoken version of a name should be deduced from the written version, and the resulting phoneme sequence can be accurately modeled e.g. with concatenating context-dependent phoneme models. However, letter-to-phoneme sequence mappings for names in various languages are not trivial.

Ideally, as presented earlier, complete names should be used for name dialing purposes. In speaker-independent name dialing this means that users should store (type) full names for each person. However, mobile terminals quite often have a limited keypad which makes writing somewhat difficult and time-consuming. Therefore, users are not keen to write full names for each person they add to their phone directory. Instead of "John Smith" one may write "John" or "John S". This approach is not well suited for speech recognition purposes since part of the important information is not given at all. If users are required to type in full names, then the information-storing phase requires more effort and many users may not be happy with that.

One small problem is also the arbitrary word order. At least advanced users often store names so that they always start with the last name, e.g. "Smith John". This may be due to the fact that they perform searches based on the last name, and that name lists are more natural to browse in the alphabetical order of the last names than the first names. However, most of these users are likely to speak the names in the reverse, more natural order, starting with the given name (naturally depending on local culture and habits). As a result, speech recognizers may have to recognize names spoken in any order, which adds to the complexity of the system.

Letter-to-phoneme sequence mapping is not trivial in a low-resource mobile terminal environment. The most straightforward approach would be to use a pronunciation lexicon including all possible names in the world. This is clearly not very practical, and more practical methods are required. In practice, either a lexicon compression method or a compact

universal pronunciation model is needed. For example, lexicon compression method proposed in [2] provides very high string accuracy with typical English names with memory consumption less than 200 Kbytes.

One additional challenge in letter-to-phoneme sequence mapping is the multi-lingual aspect. For example, in order to know how the name "David" is pronounced, one should know whether the person is French, English, or something else. This problem can be solved by a language identification approach in which at least the full written name, phone number, and perhaps the default language of the device itself are utilized.

In an ideal case, foreign pronunciation rules for names should also be utilized. For example, native English users tend to pronounce foreign names in an English way, and if that information is utilized, improved recognition accuracy is achieved.

Network based solution. Network based name dialing services that utilize speaker-independent technology and enable users to call to a pre-defined group of persons (e.g. employees of a corporation) by speaking their names are becoming increasingly common. These kinds of services have several significant advantages over the speaker-dependent name dialing approach. First, the users do not have to train the names, but the names are stored centrally. All users have the same ready-made name list. Second, these systems support all the names within a given group, and thus, users do not have to remember which names are supported since they all are. That is, out-of-vocabulary input becomes a smaller problem. Third, this service is available from everywhere, from all phones, and thus, the availability of the service is high. Fourth, this kind of service saves time in a dialing phase. In our tests, call initiation with a name dialer system was measured to take 5 sec in average. Call initiation by keying in a 10-digit telephone number takes slightly longer time, if the user remembers the number by heart. If that is not the case, it takes about 25 seconds to find an appropriate web page with the corporate telephone directory, to type in the name of the called party, to retrieve the number from the directory, and to key it in with the phone and make a call.

Based on our experience, a system that enables ready-to-use name dialing, with a sufficient amount of names, 'hooks the users' rather effectively. From the end-user point of view, it is usually of secondary importance whether the service is implemented in a mobile terminal or as a network offering.

8. CONCLUSIONS

In this paper, we discussed the usefulness of the name dialing, feature that is currently widely available as a terminal or network based offering. Based on experiments, we have observed various concerns of end-users which significantly limit the general usefulness of name dialing, making it a function which is used only occasionally, rather than regularly. Since many of these concerns are related to our 'social behavior' – rather than to the underlying technology – we know that the time will alleviate the hindering factors of the wider usage and acceptance of name dialing. In addition, we observed the potential in speaker-independent approach. Ready-to-use

technology has major advantages over systems in which extra training efforts are required from the users.

In the end, we would like to note that an important usefulness aspect of name dialing is the fact that it is bringing ASR into every pocket, and thus reducing the hindering social factors in the wide usage of ASR in our everyday lives.

REFERENCES

- [1] Dobrin C., Boda P, Laurila K. "On Usability of Name Dialing", *Proc. of Automatic Speech Recognition and Understanding Workshop*, Keystone, Dec, 1999.
- [2] Pagel V., Lenzo K., Black A. "Letter to Sound Rules for Accented Lexicon Compression", *Proc. of International Conference on Spoken Language Processing*, Sydney, Dec, 1998.