Reprinted from

SPEECH
COMMUNICATION

Speech Communication 17 (1995) 227–234

Directory assistance automation in Bell Canada: Trial results

Matthew Lennig [*,a], Greg. Bielby [a], Julie Massicotte [b]

[a] *Bell-Northern Research, 16 Place du Commerce, Montreal, Canada H3E 1H6*
[b] *Bell SYGMA, 671 de la Gauchetière Ouest, Montreal, Canada H3B 2M8*

Received 7 June 1995

ELSEVIER

ELSEVIER

Speech Communication 17 (1995) 227–234

SPEECH COMMUNICATION

Directory assistance automation in Bell Canada: Trial results

Matthew Lennig [*,a], Greg. Bielby [a], Julie Massicotte [b]

[a] *Bell-Northern Research, 16 Place du Commerce, Montreal, Canada H3E 1H6*
[b] *Bell SYGMA, 671 de la Gauchetière Ouest, Montreal, Canada H3B 2M8*

Received 7 June 1995

Abstract

Speech recognition was used to automate directory assistance in a 6-month trial with Bell Canada's public customers. The bilingual application gave the caller a choice at the beginning of the dialog to continue in English or French. Over 89% of calls were either partially or fully automated. Customer and operator reactions to the system were positive. Bell-Northern Research's flexible vocabulary recognizer, using a vocabulary of 1,700 city names and synonyms, performed well under real world conditions. Economically significant operator work time savings were demonstrated.

Zusammenfassung

Für einen mit den den Bell-Canada-Kunden durchgeführten sechs-monatigen Test zur Automatisierung der Telefonauskunft wurde ein Spracherkennungssystem angewandt. Dieser zweisprachige Telefondienst ließ dem Benutzer die Wahl, sich in Englisch oder Französisch auszudrücken. Mehr als 89% der Anrufe waren teilweise oder vollständig automatisiert. Die Reaktionen der Benutzer und Telefonisten auf den Telefondienst waren positiv. Das flexible Spracherkennungssystem von Bell-Northern Research mit einem Vokabular von 1700 Städtenamen und Synonymen lieferte dabei gute Ergebnisse. Aus wirtschaftlicher Sicht wurde eine erhebliche Reduzierung der Arbeitszeit der Telefonisten erzielt.

Résumé

La reconnaissance de parole a été utilisée pour automatiser le service des renseignements annuaire lors d'une expérimentation de six mois avec les usagers résidentiels de Bell Canada. Ce service bilingue donnait à l'usager le choix de s'exprimer en anglais ou en français. Plus de 89% des appels ont été partiellement ou complètement automatisés. Les réactions des utilisateurs et des téléphonistes vis à vis du service ont été positives. Le système de reconnaissance à vocabulaire flexible de Bell Northern Research a fourni de bonnes performances avec un vocabulaire de 1700 noms de ville et synonymes. On a pu mettre en évidence une réduction significative, du point de vue économique, du temps de travail des opératrices.

Keywords: Speech recognition; Telephone; Directory assistance; Automation; Flexible vocabulary recognition; FVR; Speaker independent; Phoneme

* Corresponding author. E-mail: lennig@bnr.ca.

Elsevier Science B.V.
SSDI 0167-6393(95)00024-0

1. Introduction

In the late 1980s, Bell-Northern Research in Montreal developed speech recognition technology for automating collect and third-party-billed calls. This resulted in the first large scale application of speech recognition in telecommunications: Automated Alternate Billing Service, introduced by Northern Telecom in Michigan Bell in May 1989 (Lennig, 1990). Although this satisfied the need to automate toll calls and is now saving hundreds of millions of dollars annually for North American telephone companies, it failed to address the other major function of operator services: directory assistance. It has long been thought that directory assistance is too difficult a task to automate with speech recognition technology.

Although there are significant technological hurdles to overcome, the rewards of automating directory assistance are substantial. Every year, telephone companies in the U.S. spend over $1.5B providing directory assistance service. It typically takes the operator about 25 seconds to complete a directory assistance call. A reduction of one second in this average work time represents a savings of over $60M a year. This paper describes a trial, conducted with Bell Canada, whose purpose was to investigate the potential for automating directory assistance using speech recognition.

The trial, dubbed Automated Bilingual Directory Assistance (ABDA), was conducted from March through August 1993 in Hull, a city adjacent to Ottawa on the Quebec side of the Quebec/Ontario border. There were 65,000 Bell Canada subscribers participating in the trial, each of whom received a letter explaining the trial prior to its start. The system either fully or partially automated 89% of the 265,805 calls received during the ABDA trial.

The trial attempted to automate language selection and requested locality. While the language selection task expected a core vocabulary of the two words *français* and *English*, the locality recognition task had a vocabulary of 1,200 distinct locality names plus 500 synonyms for those names per language. For both tasks, the recognizer performed basic word spotting, so requests such as "*Ahh, I think it's in Saint Jean sur Richelieu, Quebec*" were supported. In the final stages of the ABDA trial, recognition of listing names was introduced, and a small number of calls were automated from end to end with no operator intervention.

2. Goals of the trial

The goals of the trial were to assess the potential of speech recognition to automate directory assistance and similar services; to evaluate the potential savings of partial automation; to evaluate customers' reactions to the new technology and the effects of different user interfaces; and, for partially automated calls, to evaluate the operator interface and operator acceptance.

3. User interface

The following is an example dialog for a partially automated call. The ABDA system's speech is shown in italics, while that of the caller is indented. The caller dials 411 and hears

*Ici l'assistance annuaire Bell. Pour le service en français, dites **français**; for service in English, say **English**.*

English.

For what city?

Hull.

What is the name?

Dennis Farney.

Please hold for operator assistance.

The initial prompt greets the caller and requests that he select the language of interaction. If French had been selected, the rest of the dialog would have continued in French. Since the Hull region is predominantly French speaking, about 90% of the callers selected French.

4. Phases of the trial

The trial proceeded in phases. At each phase, more automation was introduced. In phase 1,

speech recognition was used only to determine the language. The locality name and listing name were simply recorded. At the end of the dialog, the call was queued for an operator.

When the call arrived at the operator's position, she heard a call arrival tone, the locality name, the listing name, and a different tone indicating that she was now connected to the caller. In the above example, she would hear ⟨*BEEP*⟩ *Hull. Dennis Farney.* ⟨*beep*⟩. The language indicator on the screen would show that the caller had selected English.

In some cases, based simply on hearing the recorded locality and listing names, the operator was able to complete the search without interacting with the customer. Once the correct listing was found, she simply released the call to the audio response unit that played back the requested telephone number in the language selected by the caller. For more complex queries, the operator needed to interact with the caller to determine which listing was required before releasing the call to the audio response unit.

In phase 2 of the ABDA trial, speech recognition was used to recognize the locality name. The recognized locality name appeared on the operator's screen in the appropriate field when the call arrived at the operator position. The cursor was positioned in the listing name field and only the listing name was played back. In the example, the operator would hear ⟨*BEEP*⟩ *Dennis Farney.* ⟨*beep*⟩. All the operator needed to do was to type "Dennis Farney" and press the search key.

Sometimes callers request cities outside the area code served by the directory assistance office. In phase 3 of the ABDA trial, the system redirected such calls by giving dialing instructions to reach the directory assistance office in another area code. For example, if a caller specified Toronto as the city, the dialog would go like this:

Ici l'assistance annuaire Bell. Pour le service en français, dites ***français***; *for service in English, say* ***English***.
English.
For what city?
Toronto.
For directory assistance in this city, please hang up and dial 1-416-555-1212. I repeat. For directory assistance in this city, please hang up and dial 1-416-555-1212.

In phase 4, a search type prompt was activated to distinguish residential listing requests from others. After the locality question and before the listing question, ABDA asked the caller "*Is it for a residential listing?*" In the affirmative case, the "Res" indicator appeared on the screen when the call arrived at the operator position.

In phase 5, a new prompt was added to the dialog for residential requests only, asking for the street name. The listing and street names were played back. The phase 5 dialog looked like this:

Ici l'assistance annuaire Bell. Pour le service en français, dites ***français***; *for service in English, say* ***English***.
English.
For what city?
Hull.
Do you want a residential listing?
Yes.
For what name?
Dennis Farney.
What is the street name?
King Street.
Please hold for operator assistance.

When the call arrived at the operator, the language, search type, and locality fields were filled in on the operator's screen. The listing name and street name were played back: ⟨*BEEP*⟩ *Dennis Farney. King Street.* ⟨*beep*⟩.

In phase 6 speech recognition was performed on the listing name for non-residential queries. From data collected during the previous phases, the most frequently requested business names were identified. If the recognition was successful, the requested business name appeared on the operator's screen in the listing name field when the call arrived at the operator's position. No playback was required. The operator heard ⟨*BEEP*⟩⟨*beep*⟩. She then pressed the search key and, if the search resulted in a unique number, released the call to audio response.

The last phase of the trial, phase 7, lasted only a few days. It consisted of the complete automation of a small number of frequently requested listings. ABDA recognized the language, locality, search type, and listing; it performed a database search automatically; if the listing was unique, the call was released directly to the audio response unit without operator intervention. One call sounded like this:

Ici l'assistance annuaire Bell. Pour le service en français, dites ***français***; *for service in English, say* ***English***.
English.
For what city?
Ottawa.
Do you want a residential listing?
No.
For what name?
OC Transpo.
The number is 741-4390. I repeat, the number is 741-4390.

5. Dialog modifications

The quality of users' responses depends on the wording of the prompts. Well designed prompts also make the system easier to use. The following prompt changes were implemented during the course of the trial.

The greeting prompt was changed five weeks into the trial from *"Ici Bell Canada."* to *"Ici l'Assistance-annuaire Bell."* because of the high percentage of hangups. When the prompt was changed, hangups dropped from 9.6% to 7.7%.

When we activated the search type prompt, *"Do you want a residential listing?"* it lead to a variety of unexpected answers. After a week we changed it to *"Please answer the question yes or no. Do you want a residential listing?"* This elicited higher compliance.

The French version of the listing name prompt was initially *"Quel est le nom?"* (What is the name?) but many customers answered with their own names! Therefore, eight weeks into the trial, this prompt was changed to *"Quel est le nom demandé?"* (What is the requested name?). This had the desired effect.

The English version of this prompt was also changed from "What is the name?" to "For what name?". There was no observed customer reaction to this change. The purpose of this change was to shorten the prompt, leading to a snappier, more pleasing dialog.

The dialogs used in the trial did *not* use a beep tone to indicate when the user could speak. Beep tones detract from our goal of creating a quickly paced, natural flowing dialog. They also invite users' speaking before the beep, which generates spoke-too-soon errors.

We did not employ talkthrough in the trial. Because the prompts were carefully designed and short, users' speaking before the end of the prompt was never a problem except during the language selection prompt, which was rather long and bilingual. This is the one place where we feel talkthrough would have been of benefit.

6. Customer acceptance

A customer survey was conducted by Le Groupe Léger & Léger, an external consultant. The 207 customers interviewed had used the system within the last 24 hours. The ABDA system was well accepted by customers: 93% of those interviewed stated that the system was easy to use on their first attempt. Furthermore, 99% of those interviewed said they followed all the steps from beginning to end when they first used the system and 87% were satisfied with the way in which their first call was handled.

Given the choice between the usual directory assistance system and ABDA, 41% preferred the usual service, 44% preferred ABDA, and 15% had no preference. Of the respondents who preferred the usual service, 55% said they preferred to speak with a real person and not a machine. As well, 24% of these users found the usual service faster, whereas 13% had the impression that they receive more information with the usual service. As for those who preferred ABDA, 52% found it faster and 34% easier to use.

62% of all respondents thought that ABDA was an improvement over the usual service, while 34% thought otherwise. Respondents felt that speed and practicality make the ABDA system an improvement, while those who were of the opposite view found the ABDA procedure too long and the system impersonal.

7. Speech technology

ABDA used Bell-Northern Research's flexible vocabulary recognition (FVR), an earlier version of which was described in (Lennig et al., 1992). In the current version of FVR each phoneme is represented by multiple context-dependent allophonic hidden Markov models with continuous mixture densities. The number of states used in a model varied as a function of the phoneme. The average number was four. To model English, 132 allophones were used. For French, 238 were used. Features used in ABDA consisted of mel-frequency cepstral coefficients and delta coefficients computed every 12.75 milliseconds on a 25.6 ms hamming window. An endpointer was used to determine the beginning and end of the input token.

FVR, accompanied by other interactive speech technologies, is implemented on Northern Telecom's Network Applications Vehicle (NAV), a prototype of which was used in the ABDA trial. In addition to FVR, the NAV supports digital speech recording, digital speech playback, speech compression, dual-tone multifrequency (DTMF) reception, and text-to-speech synthesis. It also supports the T-1 digital telephony interface as well as 802.3 and X.25 data interfaces. More details about NAV can be found in (Lennig and Sharp, 1993).

8. Recognition accuracy

Too often, experimenters eliminate ill-behaved user responses (tokens) from the evaluation sample. Reasons given are that these tokens do not meet a subjective reasonableness criterion, as judged by the experimenter, or that the speaker is a child or a goat. We disagree with such practices when evaluating real-world applications.

Our results are reported on 100% of the tokens. For each token, four outcomes are possible (Lennig, 1990):

- Correct Acceptance (CA): The recognizer returns a result consistent with the intent of the user. For example, if the user says any of *yes, yeah, yup, yuh, yes please*, or *yes I will* and the recognizer returns 'yes'.
- Correct Rejection (CR): The recognizer rejects a token outside the recognizer's vocabulary (an *imposter*). For example, if the user says *yesterday* in response to a yes/no question and the system correctly rejects the input.
- False Acceptance (FA): The recognizer incorrectly accepts either an in-vocabulary token or an imposter. For example, if the system accepts a *yesterday* token as the word 'yes', or accepts the city name 'Laval' when the caller said *Lasalle*.
- False Rejection (FR): The recognizer incorrectly rejects an in-vocabulary token by itself or in a word-spotting context. For example, if the system rejects the token *Ahh, I think it's in Saint Jean sur Richelieu, Quebec* during locality recognition and 'Saint Jean sur Richelieu' is in the vocabulary it would be counted as a false rejection.

The sum of CA + CR + FA + FR = 100% of the tokens. The endpointer is considered part of the recognizer: if the recognizer makes a mistake because of incorrect endpointing, this counts as a recognition error.

Language selection accuracy during the ABDA trial was as follows: 89.1% CA, 8.0% CR, 1.2% FA, 1.7% FR. While the combined rate of FA and FR is 2.9%, it is incorrect to equate this result to a 2.9% error rate with forced choice (no rejection) on in-vocabulary data. The result with rejection is much more difficult to achieve.

On the 1,200 locality plus 500 synonym vocabulary, results improved steadily during the course of the trial. The average results were 81.8% CA, 8.6% CR, 1.9% FA, 7.7% FR. Towards the end of the trial, the accuracy had improved to 83.1% CA, 8.9% CR, 1.0% FA, 7.0% FR. The imposter rate for the locality question was 6.5%.

9. Office average work time

The purpose of automating directory assistance is to save operator time. In the operator services industry, performance is traditionally measured by the average time it takes an operator to handle a call. This is called average work time (AWT). This measure can either be computed for an individual operator or for an entire office or telephone company.

AWT is computed over a given interval of time as the total operator work time divided by the total number of calls handled. Work time is the aggregate time (during the interval of interest) that the operator is actually working on calls, excluding her idle time between calls. Board time is the amount of time an operator spends plugged into her position, either working on a call or waiting for the next call to arrive. Board time = work time + idle time. Occupancy is the ratio of work time to board time and normally runs in the 92% to 97% range for a reasonable size operator force.

The traditional AWT measure is adequate to measure the savings resulting from partially automated calls. However, because its denominator is the number of calls handled by an operator or group of operators, it will not register any savings for fully automated calls that never reach an operator. In fact, the measured AWT may increase, since it is generally the simpler calls that can be fully automated.

We propose a generalization of AWT called office average work time (OWT). OWT can only be computed over the entire group of operators (the office) which is handling traffic subject to automation. The office defined for purposes of OWT may or may not correspond to a physical operator office, and may span geographic regions. OWT is then defined over some time interval as the ratio of aggregate operator work time for the office to the number of calls handled by the office (including those fully automated).

10. Operator acceptance

The general feeling among operators about ABDA was positive. They liked not having to talk as much, repeating the same things over and over. They found the system easier on their voices (fewer trips to the water fountain). There was also less keying required, which was especially appreciated by operators with physical disabilities.

The major negative comment was that operators found they were not busy enough. This is due to the low occupancy of 40%, dictated by the small size of the ABDA operator pool used in the trial (4 positions) and the need to avoid excessive queuing time.

11. Experimental design

Twenty-six operators participated in a controlled experiment to measure work time savings. The operators worked part of the time on ABDA positions and part of the time on control positions which were not connected to ABDA but which were configured to receive only 411 calls.

The control task was imperfect in two respects. First, traffic handled on the ABDA positions originated only from the 65,000 lines selected in Hull and the surrounding region, while the control positions received 411 traffic from a wider geographic area. More importantly, the occupancy of the ABDA positions was only about 40% compared to the control positions, which were running at 72% occupancy. AWT is thought to have a strong negative correlation with occupancy. Since it is hard to quantify, no correction has been made in the data presented below to account for this phenomenon, so the reported savings may be understated.

Another problem with the experimental setup was caused by a software upgrade installed a few weeks prior to trial start which reduced the search response time by 0.1 seconds on the control positions but which could not be implemented on the ABDA positions. To correct for this effect, it is necessary to multiply the average number of searches per call by 0.1 seconds and subtract the resulting time from the AWT. This correction, which amounts to 0.3 s for residential calls and 0.2 s for non-residential calls, has been made in the data presented below.

12. Observed savings

These results represent long term averages over the trial. As operators and users became more familiar with ABDA, savings improved.

Directory assistance calls for residential listings represent approximately 20% of all calls. No savings were observed for these calls. We speculate that the lack of observed savings for residential queries was due to the lower occupancy in the ABDA group as compared to the control group. Adding the street name prompt (phase 5) surprisingly had no effect on AWT. This was because only about half of the callers knew the street name. The other half answered the street name question with "I don't know" or an equivalent phrase. Playback of this phrase counteracted the AWT gained when this token contained valid information. We proposed a solution to this problem but did not implement it. This proposal was to use the recognizer to listen for "I don't know" and its equivalents and, if detected, inhibit playback. We may try this in the future.

Calls for business listings make up about 70% of all calls. These calls showed an AWT saving of 3.1 seconds.

Calls for government listings make up roughly 10% of the traffic. These calls showed an AWT saving of 7.3 seconds.

Miscellaneous calls are those calls which do not result in any search being attempted, for example when someone dials 411 to ask for the time of day. In Bell Canada these calls are not billed, but take up valuable operator time. Such calls were reduced from 11.3% in the control group to 4.6% in the ABDA group, representing a significant savings, on the order of one second of OWT. The presence of an automated system discouraged people from calling 411 to get the time of day or the recipe for boeuf bourgignon.

In addition, 1.1% of all calls were redirected to directory assistance in another area code. The operator time required to handle these calls was reduced to zero, resulting in an OWT savings of roughly 0.2 s.

A small number of frequently requested business names were also fully automated. Because this capability was only implemented in the last few days of the trial, however, the number of calls automated in this way was too small to impact savings.

13. Automation of frequently requested listings

Although the focus of the Bell Canada trial was on locality automation, we also used the trial to study the potential for frequently requested listing (FRL) automation. In a study of 108,980 calls between 14 March 1993 and 11 June 1993, we found there were 30,580 unique released numbers. Fig. 1 shows the percentage of calls covered by the N most frequently requested numbers as a function of N, for both residential and nonresidential numbers. We see from Fig. 1, for example, that the 1,000 most frequently released numbers covered over 43% of the call volume and that half of the total call volume is covered by the 1,600 most frequently released numbers.

This result is encouraging, but some caution is required in interpreting it. One telephone number may correspond to several listings. Therefore, the vocabulary needed to automate the 1,000

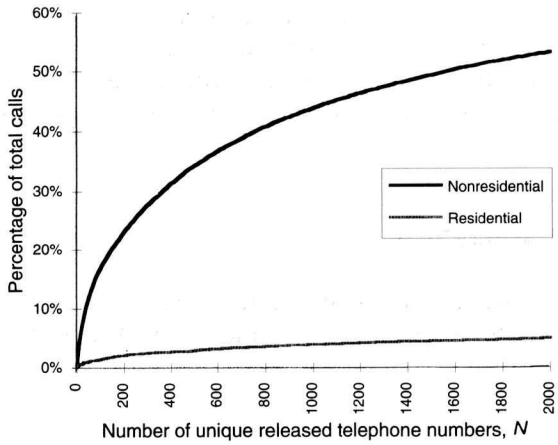

Fig. 1. Percentage of all directory assistance calls covered by the N most frequently requested telephone numbers as a function of N, shown for nonresidential and residential requested telephone numbers. Based on a sample of 108,980 directory assistance calls processed between 14 March 1993 and 11 June 1993.

most frequently requested numbers may be many times larger than 1,000. Furthermore, the most frequently requested numbers often occur in caption sets, that is, hierarchical sublistings within larger listings. Caption set automation involves more complex dialogs than automation of straight line listings. Finally, these statistics are derived from a single set of data used both to determine the N most frequently requested numbers and to estimate their coverge. From other experiments, we have observed that this within-set method biases our coverage estimates by about +5%.

In a subsequent directory assistance trial held with U S WEST in Denver in 1994, more emphasis was put on FRL automation, particularly in the later part of the trial. By the end of the trial we were fully automating 3% of the total directory assistance traffic. This was achieved using a dictionary of under 1,000 listings. Of these, less than 500 listings were actually used in fully automated calls.

14. Conclusions

We have demonstrated that directory assistance automation using speech recognition is feasible and has the potential to save significant amounts of operator work time. Most users and operators reacted positively to the technology. Two types of calls were fully automated: call redirections and frequently requested listings. This represents a world's first for speech technology.

15. Future work

BNR Montreal has developed a Northern Telecom product based on the ABDA trial called Automated Directory Assistance Service Plus (ADAS Plus). Phase 1 of ADAS Plus, which automates language selection, locality recognition, search type, and call redirection, was introduced in Bell Canada in July 1995. Phase 2 of ADAS Plus, which extends these capabilities to automation of frequently requested listings, is planned for 1996.

Other applications based on this technology currently under product development at BNR Montreal include Voice Activated Dialing, Voice Activated Network Control, and Automated Business Connections. Voice Activated Dialing is a residential service that allows subscribers to maintain a repertory of phone numbers they call frequently and associate with each a spoken name. The subscriber can call any number in the list by speaking the name he has assigned to it. This service uses speaker-trained recognition. Voice Activated Network Control allows subscribers to access network features by saying the name of the feature, e.g., "Call forward." Automated Business Connections allows anyone to dial a special access code (e.g., *0) and then say the name of a subscribing business (e.g., *Blue Line Taxi*). The call is routed to the requested business.

References

M. Lennig (1990), "Putting speech recognition to work in the telephone network", *IEEE Computer*, Vol. 23, No. 8, pp. 35–41.

M. Lennig and D. Sharp (1993), "Unleashing the potential of human-to-machine communication", *Telesis*, Issue 97, pp. 22–33.

M. Lennig, D. Sharp, P. Kenny, V. Gupta and K. Precoda (1992) "Flexible vocabulary recognition of speech", *Proc. 1992 Internat. Conf. on Spoken Language Processing, Banff, 12–16 October 1992*, pp. 93–96.