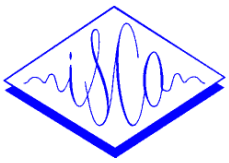

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

DIALECT-DEPENDENT SPEECH RECOGNIZERS FOR CANADIAN AND EUROPEAN FRENCH

Julie Brousseau and Sally Anne Fox

Dragon Systems Inc
320 Nevada Street
Newton, MA, 02160

ABSTRACT

In the last year Dragon Systems, Inc., a research company in the field of speech recognition, has become interested in the impact that dialect can have upon recognition rates. Research is now being done for two major dialects of the French (Canadian and European) and the English (UK and American) languages. Results from a small isolated-word speaker-adaptive speech recognizer shows that an average increase of 3% in recognition rates is obtained by a speaker dictating to a system trained with acoustic data from their native dialect. A larger system is currently under development for Canadian French and preliminary testing is going forward. These results are discussed.

10.21437/ICSLP.1992-308

I. INTRODUCTION

Dragon Systems has developed an isolated-word, large vocabulary, speaker-adaptable, Hidden Markov Model (HMM)-based speech recognition system for American English [1]. Corresponding systems for Canadian French and UK English are under development. These systems will take account of the substantial differences which exist in recognized dialects of individual languages.

It is now an established fact that Canadian French (CF) is distinct from European French (EF), just as UK English (UE) is quite distinct from American English (AE). The purpose of our research was to ascertain whether these differences are significant enough to have a substantial effect on speech recognition rates.

In this paper we describe the dialectal differences which exist between CF and EF. For the sake of illustration and comparison, parallels are drawn between dialect differences in French and dialect differences in English. Preliminary test results made on research versions of small-vocabulary recognition systems (2500 words) are presented and discussed for both French and English dialects. Then the discussion is concluded with results obtained by tests run on a larger vocabulary CF recognizer.

II. RECOGNITION SYSTEM

In DragonDictate™, the speech recognition system developed by Dragon Systems, Inc., there are three recognition components: a rapid-match algorithm, a set of acoustic word models based on HMMs, and a statistical language model.

The rapid-match algorithm [2] makes a selection of the most likely group of words, based upon a quick scan of acoustical information at the start of the word, plus word probability information from the statistical language modeling.

The acoustic parameters of an incoming word token are compared to the HMM models, and the word model which was most likely to have produced the token is determined.

The statistical language modeling is based on the frequency count of words and word n-grams. These frequency counts are obtained through analysis of a large quantity of computer-readable text from many different sources.

DragonDictate™ has been designed with a user-interface for dictation and is compatible with PC word-processor and spreadsheet software. It is built using acoustic tokens from one speaker, the "reference speaker". Using these base models, the system can then "learn" the voice patterns of a new speaker by a process of adaptation. This involves three processes: rapid-match adaption, Hidden Markov Model adaptation, and language model adaptation. This adaptation mainly takes place during the first 2000 words of using the system. Later, the system continues to adapt but at a slower rate.

III. DIALECT DIFFERENCES

The major differences which can occur between two dialects are apparent in their phonological, phonetic and lexical composition.

3.1 Phonology

CF has a marked phonological difference in the /ɛ/ sound [3] which is not produced in EF. The difference is as a result of the duration and the timbre of this vowel and can be witnessed in words such as:

Canadian French /ɛ/ ≠ /ɜ/		**European French** /ɛ/		
"fête"	[fɜt]	"fête"	[fɛt]	*(birthday)*
"faites"	[fɛt]	"faites"	[fɛt]	*(do)*
"maître"	[mɜtr]	"maître"	[mɛtr]	*(master)*
"mettre"	[mɛtr]	"mettre"	[mɛtr]	*(to put)*

That is, "fête-faites" and "maître-mettre" are homonym pairs in EF but not CF. The alternate /ɛ/-like vowel in CF, called /ɜ/, has a timbre intermediate between /ɛ/ and /e/ and is often diphthongized.

A similar phenomenon can be observed in UE in the /iə/ sound. Take note of the examples listed below.

UK English	**American English**
"career" -> [kə/'riə]	"career" -> [kə/'rir]
"Korea" -> [kɔ/'ri/ə]	"Korea" -> [kɔ/'ri/ə]

That is, "career" and "Korea" are homonyms (or near-homonyms) in UE but not in "general" AE. In this case it is easy to understand how vowel differentiation can have a noticeable effect on the recognition rate of a distinct dialect, because the different sounds distinguish different words, hence leading to a situation of fewer confusable words in one dialect as compared to the other.

3.2 Phonetics

Aside from the phonological issues there are audible phonetic distinctions which occur in CF speech when compared to EF. Some of these distinctions are contextually-based. Two good examples to illustrate this are vowel laxing and assibilation.

In CF the high vowels /i, y, u/ are laxed and become [I, Y, U] in a closed final syllable if the closing consonant is not a

lengthening one [4]. More formally:

/i,y,u/ -> [I,Y,U] / __C {except /r,v,z,ʒ/}

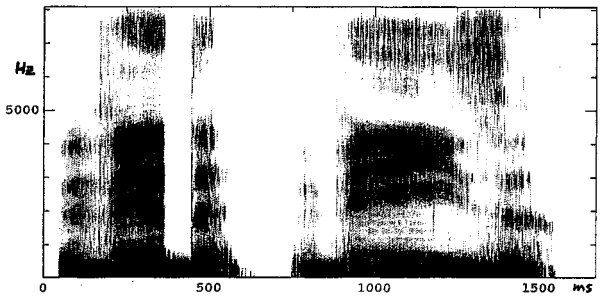

Fig. 1 Wideband spectral representation of the CF pronunciation of the words "vide" and "vise".

Figure 1 shows the specific acoustical behavior of the vowel /i/ in the words "vide" and "vise" pronounced by a CF speaker. In addition to the change of position of F1 and F2 of /i/ between these two words, the spectral representation also clearly indicates that the vowel duration is shortened when the /i/ is laxed.

Similarly, the apical stops /t,d/ usually become the affricates [ts, dz] when they are followed by /i,y,j,ɥ/. That gives the rule:

/t,d,/ -> [ts, dz] / ___ /i,y,j,ɥ/

This phenomenon is called assibilation and can be observed on a wide band spectrogram by the presence of either one of the fricatives /s/ or /z/ following the stop /t/ or /d/ (Figure 2).

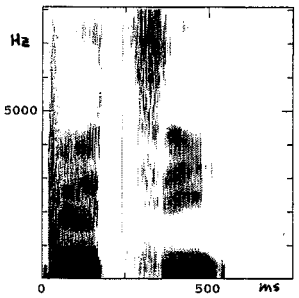

Fig. 2 Wideband spectral representation of the CF pronunciation of the word "petit".

The two spectral differences observed in Figure 1 and Figure 2 above do not appear in EF and underline the fact that such contextual behavior can lead to misrecognition if the recognizer does not take into account the phonetic differences that may occur between dialects.

Other phonetic differences are context-independent. In CF, clear distinctions are still made in the oral vowel pair /ɛ/-/e/ and the nasal vowel pair /ɛ̃/-/œ̃/. In EF, there is a strong tendency to merge the /ɛ/ sound and the /e/ sound into /e/ [5], and the /ɛ̃/ and /œ̃/ sounds into /ɛ̃/. Take for example words such as:

Sound	Examples	Canadian French		European French
/e/	"j'aurai"	[ʒɔre]	⇒	[ʒɔre]
/ɛ/	"j'aurais"	[ʒɔrɛ]	⇒	[ʒɔre]
/ɛ̃/	"brin"	[brɛ̃]	⇒	[brɛ̃]
/œ̃/	"brun"	[brœ̃]	⇒	[brɛ̃]

Similarly if the major phonetic differences found in UE and AE are carefully considered, the case for misrecognition of words as a result of dialect become evident. There are vowel sounds in UE which are not used in AE pronunciation, such as

Sound	Examples
/əʊ/	"goat", "show"
/ɒ/	"odd", "pot"
/ɛə/	"square", "various"
/juə/	"cure", "jury"
/ɜ:/	"nurse", "stir"

Also AE and UE have a great divergence of pronunciation concerning the "r"sound. In UE pronunciation, the post-vocalic "r" sound following the vowel is mostly lacking, for example in words such as:

Pronounciation	Examples
[hɜ:t]	"hurt"
[fɑ:]	"far"

Whereas the orthography of the word is the only indication that the "r" was once pronounced in UE, there is a clearly audible postvocalic "r" in AE.

3.3 Vocabulary

There can be substantial differences in vocabulary between dialects. In building a speech recognizer, this vocabulary difference is important for the statistical language model because particular words will occur more frequently in one dialect than the other. Changing the frequency occurence of a word, in order to make a better representation of the vocabulary of a given dialect, offers a higher probability that the correct word will be recognized. Listed below are examples to illustrate some vocabulary differences which occur between CF/EF and UE/AE.

Canadian French	European French	
"chaudron"	"casserole"	*(casserole)*
"fournaise"	"chaudière"	*(furnace)*
"chum"	"copain"	*(boyfriend)*
"blonde"	"copine"	*(girlfriend)*
"souper"	"dîner"	*(to have dinner)*

UK English	American English
"lift"	"elevator"
"estate-car"	"station-wagon"
"porridge"	"oatmeal"
"courgette"	"zucchini"
"draughts"	"checkers"
"aeroplane"	"airplane"
"washeteria"	"laundromat"

3.4 Punctuation and Spelling

While punctuation and spelling differences do not have any impact on recognition rates, they are important if a vocabulary is to be customized for a distinct set of users. In CF one has the use of capitalized accented characters, e.g., "Île-du-Prince-Édouard" and "Nouvelle-Écosse". In UE you have the word "full-stop" in place of the AE "period", "inverted commas" in addition to "quotes". Examples of spelling differences between UE and AE include such pairs as colour/color, encyclopaedia/encyclopedia and licence/license.

III. EXPERIMENTAL PROCEDURE

To test the effect of difference in dialect in French, our procedure was first to build two small research systems (2500 words), one using CF base models, the other using EF base models. The words were taken from the 27 chapters of *Le Petit Prince* by Antoine de Saint Exupéry [6]. Word frequencies for the language model, which was based only on unigrams, were also taken from the entire book. Five chapters were used for adaptation and testing (chapters 1, 2, 3, 7, and 8).

Three speakers were recorded for each dialect. The first speaker was the one who recorded the base files, the reference speaker. Speakers 2 and 3 were other native speakers of the dialect.

The tests consisted of running the tokens collected from these speakers through both systems and comparing the recognition rates obtained. Exactly parallel tests were run for UE and AE English, using small systems based on the vocabulary of the translation of *Le Petit Prince* (*The Little Prince*).

IV. RESULTS

4.1 Results for the *Little Prince* Tests

Tables 1 and 2 present the test results on approximately 2300 words for three native speakers of each dialect. Scoring began after the recognizer had adapted to 2000 words (with corrections) of a given speaker. Adaptation continued during the test on the following 2300 words mentioned above. The gender of each speaker is indicated by (F) or (M). The "R" following points out the reference speaker.

In the DragonDictate™ interface, after the user says a word, the system always shows a "Choice List" of up to nine most-likely words recognized. A word appears on the Choice List if it is determined by the system to be among the nine most likely words to have been spoken and if its likelihood exceeds a threshold. If the top choice was not the correct word, the user can choose another word off the Choice List with one voice command. If the correct word was not on the Choice List, the user may spell it by voice.

For each dialect, the three columns in the tables specify the percentage of words that were correctly recognized (column "Correct"), the percentage of words which were among the most-likely words returned by the system (column "Choice List"), and the percentage of words that were misrecognized and which did not appear on the Choice List (column "Error"). The percentages shown in the "Choice List" and "Error" columns add to 100% within round-off error.

	CF SYSTEM			EF SYSTEM		
	Correct	Choice List	Error	Correct	Choice List	Error
Speaker						
CF (F)R	88	99	0	77	95	5
CF (F)	79	97	3	76	94	5
CF (M)	83	98	2	79	94	5
EF (F)R	82	95	5	88	98	1
EF (F)	78	92	7	80	94	6
EF (F)	80	95	5	83	96	3

Table 1

	UK SYSTEM			AE SYSTEM		
	Correct	Choice List	Error	Correct	Choice List	Error
Speaker						
UK (F)R	95	99	1	90	97	3
UK (F)	91	96	4	87	94	5
UK (M)	89	96	4	82	92	7
AE (F)R	84	95	5	85	98	1
AE (F)	89	96	3	87	98	2
EE (M)	87	95	5	90	97	3

Table2

By looking at the percentages shown in Tables 1 and 2, it first appears that each reference speaker, except AE(F)R, obtains a higher "Correct" rate than the two "test speakers" of the same dialect. The better results of the reference speaker are not surprising because the recognizers were built upon the reference speaker's voice. The AE(F)R results are unexpected because they show that her "Correct" rates are 2% and 5% below AE(F) and AE(M) respectively. However when the "Choice List" rates obtained by the AE reference speaker are considered, she does as well as the female test speaker and 1% better than the male test speaker and therefore obtains a lower percentage "Error" than either test speaker.

It is mainly interesting to look at the results of each non-reference speaker across dialect. In most cases, except for AE(F), each speaker gets a higher "Correct" recognition rate when talking to a recognizer built with acoustic models of their native dialect. This difference in recognition rates starts at 2% for EF(F) and reaches 6% for UK(M). The lowest "Correct" percentage for a native speaker on a "native system" is obtained by CF(F) (79%), but she obtains 97% as a "Choice List" rate and a 3% "Error" rate as compared to 94% and 5% obtained on the EF system. In fact, in all cases, the "Choice List" rates of a given speaker are higher when using a recognizer of the corresponding dialect and the "Error" rates lower. The jump that takes place between "Correct" rates and "Choice List" rates could be a result of a high percentage of homonym words. This is an especially big factor in the recognition of French.

Above all, the results shown in Table 1 and 2 tend to indicate that dialects of language can affect the recognition rates. The magnitude of the effect is similar for the two French dialects and the two English dialects studied. Averaging the data from all eight test speakers (i.e., disregarding the reference speakers) of all four dialects, we see that an average increase of 3% is obtained by a speaker dictating to a system trained with acoustic data from their native dialect. This shows the need for our approach. At the present time, we know of no other speech recognition system which explicitly takes dialect differences into account.

4.2 Preliminary results for an 18,000 word Canadian French Speech Recognizer

The results obtained on the small vocabulary recognizer led us to begin working on a 25,000-word Canadian French speech recognition system, similar to Dragon Systems' current American English recognizer known as DragonDictate™. In the near future a separate European French version is also being developed. Presently the CF speech recognizer has a vocabulary of 18,000 words and has a completely translated user interface. This allows a user to interact with the system entirely in French.

18,000-word Canadian French Recognizer				
Speaker	Correct	Choice List	Error	New
#1 CF(F)R	76	88	6	6
#2 CF(F)	75	88	6	6
#3 CF(F)	71	85	10	6
#4 CF(M)	68	80	14	6

Table3

The percentages showed in Table 3 are preliminary results obtained by four native speakers of Canadian French. The scoring began after the speaker had dictated 1735 words (with corrections). The test text consisted on the first chapter of *La détresse et l'enchantement*, a novel written by Gabrielle Roy [7]. The test text was composed of 2403 words. The first speaker is the reference speaker while the three others are test speakers. Speakers 2 and 3 are female and Speaker 4 is male. A fourth column, "New", showing the percentage of new word errors, has been added in this table.

At first sight, the percentage of "Correct" recognition rates may seem low, especially in comparison to those presented earlier in Table 1. When analysing these new results, it is important to keep in mind that they were obtained after an enlargement of the vocabulary size by 15,500 words. For the purpose of speech recognition this is an important fact because the more the vocabulary size increases, the easier it becomes to find confusable words. In French, this is especially evident with the number of homonym sets that can be found in the language such as:

"cher, chair, chère, chers, chères, chaire"
"au, aux, haut, eaux, hauts, oh, eau, ô"
"est, ait, es, aient, aie, hais"

For each non-reference speaker, the second column in Figure 3 shows new results obtained by not counting the homonym errors (i.e., word was not considered a misrecognition if one of its

homonyms was chosen by the system). By doing so, the "Correct" rates increased by 8% for Speakers 2 and 3 and by 7% for Speaker 4, to now reach recognition rates of 83%, 79% and 75%, respectively.

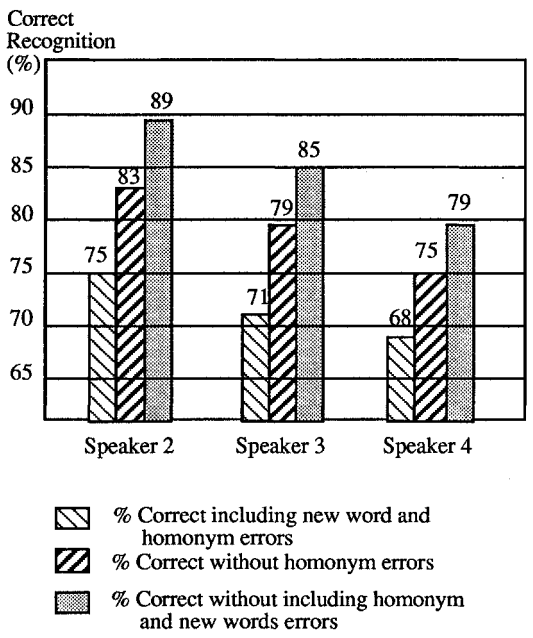

Figure 3

The problem of homonym errors can not be entirely solved by using a language model based only on unigrams. A first approach to reduce the number of homonym errors will be to create a new language model based on digram frequencies. The recognizer will then "know" that "il était" is more likely to occur than "il étais". At the moment, the language model does not include such information, and that may explain the major difference in Table 3 between the rates in the "Correct" column and in the "Choice List" column. Having a language model based on digrams won't necessarily correct all homonym errors, but a previous study with our experimental European French speech recognition system [8] showed that an increase of at least 4% of "Correct" recognition can be achieved.

Another important factor which was not relevant in the previous test (the *Little Prince* test) was new word errors. In Table 3, a fourth column ("New") showing the percentages of new word errors needed to be added. Unlike in the *Little Prince* test, the language model of this bigger system is general. The 18,000 words come from a list of the most frequent words used in a large body of diverse Canadian French texts. Unigram frequencies were also counted from these texts. Six percent (6%) of the words in the test text were not found in the vocabulary of the recognizer. Our goal for the final 25,000-word CF recognizer is a new word error rate not exceeding 5%.

New word errors are not really misrecognition errors. With DragonDictate™, a user has the possibility of adding a new word either by typing it or by spelling it by voice using the International Communication Alphabet. Once a new word has been entered by a user, it can be repeated and the system will then recognize it. We decided to leave the new words in the test text because we wanted the results to be representative of what a real user would obtain. However, to have a good representation of the recognition rate apart from vocabulary issues, it is of interest to know what the recognition rate would have been without the effect of new words. For each speaker the third column in Figure 3 indicates the percentage "Correct" obtained without the homonym and the new words errors. (A word was not considered a misrecognition if a homonym was chosen by the system, and new words were omitted from the test.) These results show that recognition rates of 89% (Speaker 2), 85% (Speaker 3) and 79% (Speaker 4) are now reached by the three test speakers. On average, this means an increase of 13% in the recognition rate.

These latest results approach those obtained by Dragon's 25,000-word American English recognizer [1]. We are encouraged by these results, especially considering the fact that our recognizer is able to achieve them with a large vocabulary, in real time, and running on a PC.

V. SUMMARY

In this paper it as been shown that Canadian French and European French are in many ways different and that these differences are comparable to those existing between UK English and American English. On average, a 3% better recognition rate was obtained on a small-vocabulary recognizer when the speakers were dictating on a natively-trained system. On a 18,000-word Canadian French System the recognition varies from 68% to 75%, but increase to a range of 75% to 89%, if homonym and new word errors are not considered. In the future we will be working on improvements, in particular on vocabulary enlargement and on improving the language model.

VI. ACKNOWLEDGEMENTS

We would like to thank *Les Publications Transcontinentales Inc* and *La société Radio-Canada* for their contribution by providing us computer-readable texts. We also thank the Royal Institute of Technology, Stockholm, for assistance with the French pronunciations.

VII. REFERENCES

[1] Bamberg P.G., "Adaptable Phoneme-Based Models for Large-Vocabulary Speech Recognition", *Proceedings of the ESCA Tutorial and Research Workshop on Speaker Characterization*, University of Edinburgh, U.K., June, 1990.

[2] Gillick L.R. and Roth R., "A Rapid Match Algorithm for Continuous Speech Recognition", *Proceedings of DARPA Speech and Natural Language Workshop*, Hidden Valley, Pennsylvania, June 1990.

[3] Santerre L., "Deux E et deux A phonologiques en français québécois", *Cahier de linguistique*, 4, 1974.

[4] Walker D.C., *The Pronunciation of Canadian French*, University of Ottawa Press, Ottawa, 1984.

[5] Callamand M., *Méthodologie de l'enseignement de la prononciation - Organisation de la matière phonique du français et correction phonétique*, CLE international, Paris, 1981.

[6] Saint Exupéry A.de, *Le Petit Prince*, Harcourt Brace Jovanovich, New York, 1943, 113 pages.

[7] Roy G., *La détresse et l'enchantement*, Boréal, Montréal, 1984.

[8] Bamberg P. et al., "Incorporating natural-language information into a stastical language model for large-vocabulary recognition" *Proceedings of Speech Tech*, New York, 1992.