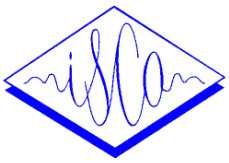

2nd European Conference on
Speech Communication and Technology
EUROSPEECH '91
Genova, Italy, September 24-26, 1991

COMPUTER-AIDED, VOICE-BASED, MEDICAL REPORT PREPARATION: AN APPLICATION TO RADIOLOGY

R.Billi, P.Buttafava, P. De Stefani, M.Gamba, D.Voltolini

Olivetti Systems & Networks, Direzione Ricerca e Sviluppo,
via G.Jervis, 77 - 10015 Ivrea (TO)

ABSTRACT

This paper describes an application of speech recognition and synthesis aimed at simplifying the procedure of creating medical reports. The hardware configuration consists of a PC equipped with 2 boards, devoted respectively to speech synthesis and recognition. The system, which has been specialized for radiology, permits to dictate a report in isolated word mode, using a vocabulary which contains about 97% of the words found in a large corpus of radiological reports. To allow fast creation of reports for normal cases or for frequent pathologies, the system permits to use coded sentences which can be directly called or, alternatively, identified by means of a simple question/answering procedure.

10.21437/Eurospeech.1991-230

1. Introduction

The availability of speech recognition and synthesis at an accessible cost and as add-on boards for personal computers opens a wide range of applications and promises to significantly increase the market for speech technology.

The most appealing applications are those in which speech really offers an advantage over alternative conventional or emerging input/output devices. That is, for instance, when the hands/eyes are busy, the operator must have freedom of movements, and so on.

Medicine has since long been recognized as one of the fields which has these favourable conditions, particularly for speech recognition.

To check the actual potential for applications, we have started the development of a PC-based work-station for dictating radiological reports. This choice comes from the observation that a real need exists of improving the current procedures for transcribing radiological reports and that â large amount of reports is actually produced each year.

When a radiologist analyzes radiographs, his hands and his eyes are busy. Consequently the best he can do now is to record his report on a tape. Later, a typist transcribes the text on paper. This is a time consuming procedure.

We have built an application of the technology which Olivetti has developed for text-to-speech synthesis [1] and very large vocabulary speech recognition [2] over the past years and which, for recognition, forms the basis for the on-going extension to 6 other European languages (Esprit Project *Polyglot-1*).

The main characteristics of our recognition approach are the following:

- Very large RAM-resident vocabulary, up to 10,000 words with a single recognition board and up to 60,000 words with two boards.
- Fast recognition even with 60,000 words vocabulary (< 2 sec.).
- Isolated word recognition (max throughput 50 words/minute).
- Quick training for a new speaker (independent on vocabulary).
- Easy insertion of new words in the vocabulary.
- Contextual analysis based on a statistical language model.

In order to permit fast creation of reports for normal cases or for frequent pathologies, while giving the radiologist the complete freedom for describing the complex cases, two different operating modes have been implemented, namely the guided-mode reporting and the free-text dictation, respectively illustrated in sections 2 and 3.

2. Guided-mode reporting

It is quite evident that an important percentage of all the medical reports is relative to normality or to frequent pathologies. In particular for radiology it has been measured [3] that this percentage is above 60% for almost all the kinds of reports, with particularly high percentages for thorax (78%) and mammograms (84%).

There is an interest to simplify and speed-up the compilation of the frequent cases as well as to increase the uniformity, completeness and correctness of reports. Most radiologists use coded expressions to describe frequent cases. The code identifiers can be used with word processors to quickly retrieve the associated texts.

The problem with this approach is that the physicist does not normally use word processors, nor it is easy for him to remember all possible codes.

To solve this problem we have designed an interactive question/answer vocal interface. The radiologist is led through a guided set of choices by the computer which, for each answer, selects, from a data base, the next question and eventually retrieves sentences or fragments of text, which are arranged to form a complete report.

In this way a typical report can be prepared by saying only few words and, more importantly, the radiologist is not needed to look away from the radiological image.

It must be possible to switch to free-text dictation, to modify any part of the previous text or to introduce further comments.

Another advantage of this guided procedure is that, at any time, the number of possible answers is very limited and consequently the recognition error rate is very low. In fact it is even possible to use the system, in the guided-mode reporting, without doing the training.

One of the features of our recognition system, which permits to exploit the constraints in the possible answers, is the possibility of loading many different dictionaries, activating each of them, or combinations of them, at run time.

The dialogue and the coded sentences have been designed with the help of a radiologist [3]. The grammar underlying the dialogue has a very simple structure which is described in fig.1 . All the details of the question/answering procedure can be introduced or modified by using a software tool, specifically designed for that goal.

An example of an actual dialogue is the following:

Q: kind of examination?
A: mammogram
Q: side?
A: right
Q: condition?
A: pathological
Q: kind?
A: cystic
Q: size ? (cm.)
A: two
A: end-number
Q: time to next check ? (years)
A: one

Note that the questions are pronounced by a text-to-speech system, thus not requiring to look to the screen.

After this short dialogue a complete report (6 lines of text) is produced and printed, ready for signature.

3. A statistical language model for free-text dictation

The goal of the language model is to select, among the word candidates proposed by the acoustic recognizer, the most likely word sequence. The information available to the language model is grammatical and statistical.

Our language model is essentially that described in [2], i.e. a Markov model based on bigrams of grammatical tags. The statistics used at grammatical level proved to be sufficiently general and did not require to be re-estimated for each specific linguistic domain. We have estimated them by using a collection of different texts totalling 550,000 words.

The corpus problem arose when we faced the problem of defining the vocabulary to be used for the dictation of radiological reports. In fact we were not able to collect, from the radiologists, a word list that had sufficient coverage on the reports, mainly because the radiologists specified only the technical words while they omitted to include very frequent words, particularly for verbs and adjectives.

The solution we adopted was to collect a large corpus and perform a frequency analysis of the used words.

We collected about 250,000 words coming from radiological reports of any kind. From our frequency analysis we found that the number of different words is 8,049. Excluding from the recognition vocabulary all the words with a single occurrence (2,549), we obtained a vocabulary of 5,500 words, covering the 97% of the corpus.

These results are reasonably similar to those published by IBM for the Italian language [4], considering the different corpora.

Obviously the vocabulary we got is very specific and can be used only for dictating radiological reports. For example the word "lesioni" (lesions), which occupies the position 29th in the radiological lexicon, was not found in the first 5,000 word forms of a frequency lexicon of the Italian language.

It makes little sense to have a quite general linguistic domain while the dictionary is very specific.

Thus we came to the obvious decision of introducing in the language model the individual word frequencies to improve the predictions of the language model. Using these frequencies, the language model perplexity decreased from 2032 to 209 (the perplexity is a measure of the dictionary size, which takes into account the uneven distribution of word probabilities). The reduction of the global error rate confirmed the usefulness of this choice.

A further decrease of perplexity could be obtained by partitioning the dictionary in different subsets corresponding to parts of the human body and switching across them depending on the kind of report.

To this end, we have split the corpus as follows:

1) Breast (3.02 % of the whole corpus)
2) Skull and Contents (5.25 %)
3) Face, Mastoids and Neck (4.30 %)
4) Spine and Contents (9.11 %)
5) Skeletal System (17.46 %)
6) Heart and Great Vessels (0.21 %)
7) Lung, Mediastinum and Pleura (29.95 %)
8) Gastrointestinal System (15.77 %)
9) Genitourinary System (16.81 %)
10) Vascular and Lymphatic Systems (2.11 %)

For example, using the sub-vocabulary 7, we obtained a vocabulary size of 3644 with a perplexity of 65.

To do the same for all specific domains, we need a larger corpus. However, we believe that this strategy could be fruitful.

4. Toward a work-station for radiological reporting

The application should permit the switching between the free-text dictation and guided-mode, in order to minimize the limits involved in each of these approaches while exploiting their advantages. The drawbacks and advantages of the two procedures are summarized in fig. 2 .

In a more advanced perspective, the integration of high quality digital image acquisition and display with voice-based input and data base management systems over local area networks, will permit to build a very powerful work-station which offers, in a unique system, all the functionalities needed by the radiologist.

In other words he will be able to instantly retrieve a radiographic image from the database and display it on the screen, together with the patient's personal information or the previous analyses and reports, or, alternatively, create, by using speech, a new report for the last radiographic image and store it in the database, without having to re-introduce the patient's data, which are already there.

The technology for doing all that is already available. What is really needed is the standardization of a common hardware platform and of data representation formats.

In the meantime it is necessary to experiment with speech input and to let the radiologists familiarize with this new way of producing reports.

Acknowledgements

This work could not have been done without the enthusiasm and work of Prof. A.Cugini, of the Ivrea Hospital, who designed the database for guided reporting and without the support of several Hospitals which provided the data used for statistical analyses.

References

[1] E.Vivalda, "Italian Text-to-Speech Synthesis: the Linguistic Processor", Olivetti Res. & Tech. Review, 7, 1987.

[2] R.Billi et al., "A PC-Based Very Large Vocabulary Isolated Word Speech Recognition System", Speech Communication, 9, 1990.

[3] A.Cugini, "Refertazione radiologica interattiva a viva voce con il sistema Olivetti", Comunicazione alla 1ª Conferenza Internazionale di Radiologia Digitale e PACS.

[4] A.Fusi, B.Vidal, "Il riconoscimento automatico della voce per la refertazione radiologica", La Radiologia Medica, n.81, 1991, Edizioni Minerva Medica - Torino.

SYNTACTIC TEMPLATES

S: <QUESTION> <ANSWER> <ACTION> <S> | < >

QUESTION: <SYNTHESIZED SPEECH>

ANSWER: <WORD> | <NUMBER> | <COMMAND>

WORD: any word belonging to the active word set

NUMBER: <DIGIT><NUMBER> | < >

DIGIT: <zero>|<one>|<two>|<three>|<four>|<five>|<six>|<eight>|<nine>

COMMAND: <END> | <GO_BACK> | <FREE_TEXT>

ACTION: <S_TEXT> | <FREE_TEXT> | <COMMAND EX> | <>

S_TEXT: stored text

FREE_TEXT: a text that is entered by voice

Fig.1 The grammar of guided-mode reporting

	GUIDED REPORTING	FREE-TEXT DICTATION
ADVANTAGES	- Uniformity - Rapidity - Completeness	- Generality - Freedom of expression
DRAWBACKS	- Rigidity - Limited application	- Requires concentration - All words must be uttered - Greater risk of errors

Fig. 2 Comparison between two alternative reporting modes.