A SPEECH DRIVEN LANGUAGE TRANSLATION SYSTEM

F.W.M.Stentiford and M.G.Steer*

ABSTRACT

An effective approach to the direct translation of speech between languages is presented. Keyword spotting techniques are used to overcome the inaccuracies of speech recognition and the uncertainties of natural language.

INTRODUCTION

The convergence of the technologies of speech recognition, text-to-speech synthesis and machine translation suggests the possibility of automatic speech translation (ref 1). Major problems however, have still to be solved in the handling of recognition errors and the parsing of conversational speech. Many parsers are extremely fragile in the sense that a failure in the search for potential parses is taken to mean that an incorrect path has been selected rather than an indication of erroneous input. Such parsers will yield solutions only if the input conforms precisely with the grammar and will reject input deviating even by one word (ref 2). A process which controls the dialogue can restrain the form that possible user utterances can take. However, the user normally has to be already aware of the general form of acceptable grammatical structures (ref 3,4).

Considerable effort has been applied to the processing of both spoken and textual natural language and several approaches have been investigated which are capable of negotiating the vague and fragmented nature of human dialogue. Conceptual parsers attempt to extract key ideas from the input and ignore other parts which perhaps may contain errors and omissions. They are thus immune to many errors but not to those associated with the key ideas themselves (ref 5,6).

Pattern matching is another analysis technique which matches the input against a set of patterns of words. This approach was first exemplified in ELIZA (ref 7) which appeared to cope with a wide range of human dialogue if only at a very shallow level. Again this approach is unaffected by errors occurring in parts of the input not involved in the matching process (ref 8,9). An advantage of pattern matching is its ability to handle idioms which by definition can only be recognised and interpreted as a whole.

The principal limitation of pattern matching lies in its failure to analyse much of the redundant material present in natural language. The regularities reflected by auxiliary verbs for instance, are more easily represented by a grammar than by an exhaustive list of word patterns.

This paper describes a pattern matching technique which handles dialogue regularities in a non-redundant fashion. A procedure for the extraction of information bearing key words is defined which enables a large set of phrases to be maximally distinguished from each other. This provides immunity to recognition errors even if they occur in critical parts of the utterance. At the same time it allows the identification of phrases to

* British Telecom Research Laboratories, Martlesham Heath, Ipswich, Suffolk, IP5 7RE, England.

be quite tolerant of ungrammatical and fragmented input. The set of phrases used in this work has been translated into several European languages. This together with the appropriate text-to-speech synthesisers provide the basis for a multilingual speech driven phrase book.

PROBLEMS OF SPOKEN DIALOGUE

The informal and non-grammatical nature of natural language is a fact of life (ref 9). The problems which arise are especially severe in the case of spoken dialogue which normally contains a much greater variety of unpredictable expressions than text (ref 10). The inaccuracies of speech recognisers impose an even greater handicap on any mechanism which attempts to extract meaning from a spoken utterance.

The performance of speech recognisers is influenced by the size and content of the vocabulary. In every application an acceptable compromise must be reached between the cost of handling recognition errors and the vocabulary size. In the context of speech translation no recogniser exists which is capable of handling a suitably large vocabulary at a level of accuracy which would retain the original grammatical structure for subsequent processing (ref 11).

If existing speech recognisers are to be used for the input and translation of more than a few awkwardly constructed word combinations, then such recognisers must be operated in a word spotting mode. The limited vocabulary can then be confined to a carefully selected set of keywords which effectively extract necessary information. It should then be possible to process a very wide range of utterances without the problem of very large vocabulary recognition or a heavy dependence on syntactic analysis for error recovery.

When communicating within a limited domain of discourse, it is nearly always possible to specify all the required message concepts likely to be transmitted. Such phrase books have been written for example, for air traffic communications (ref 12) and international telephone operators (ref 13). Difficulties arise when the user cannot remember the precise contents of a large phrasebook and wishes to access one of the messages using his own natural speech.

PHRASE IDENTIFICATION

The selection of keywords from the total vocabulary spanned by a phrasebook is governed by the contribution that each keyword makes towards the distinction of each phrase from all others. In this sense a word which is merely present in one phrase and found nowhere else is of less importance than a word which occurs in 50% of the phrases. Furthermore the performance of a keyword is dependent on the set of keywords already selected. In an ideal case maximum information is extracted when each keyword is present in orthogonally different binary partitions of the set of phrases. In practice this is usually not possible but it does indicate a useful criterion for keyword selection. This approach has been used to extract features for pattern recognition where it is a prime requirement that such features should act independently of each other (ref 14,15). For example, consider the three phrases :

A. Who do **you** want to **speak** to ?
B. **I** cannot hear **you**.
C. May **I** **speak** to Mr Smith please ?

The three keywords (**underlined**), "you","speak" and "I" each occur in two phrases and an optimal separation of 2 between the three phrases is

achieved. A formal specification of a keyword selection criterion is given in reference 14.

PHRASE VARIATION

A major disadvantage of phrasebooks or sentence dictionaries is their inability to cope with any variation in wording from that held in the phrasebook. This problem arises, for example, with names, places and times such as, "Is Mr. *Smith* in the office" and "Please phone back next *Tuesday*". Much of the difficulty is avoided by allowing the simultaneous recognition and temporary storage of each individual spoken word. This enables the implementation of two useful functions :

Firstly, as proper nouns are not normally recognised or translated, the original speech utterance is coded and transmitted to the receiving end for embedding in the foreign speech output.

Secondly, times and dates are recognised by implementing a two-pass recognition process. Once a phrase is correctly identified the location of the date or time within that phrase is known, or can be deduced. The speech recogniser is then loaded with new templates corresponding to the vocabulary of times and dates, and the appropriate parts of the stored speech utterance are replayed for recognition.

This two-pass recognition technique effectively increases the recognisable vocabulary of the system without degrading the performance. The sub-vocabularies may be extended to include towns, countries, the names of products or other categories of phrase.

RESULTS

A set of over 400 business letter phrases was analysed first to produce a frequency ordered list of the 1000 or so different words they contained, and then to extract a subset of 100 keywords in accordance with the phrase separation criterion. No more than a hundred keywords were chosen because of performance limitations of the isolated word speech recognition device being used.

The 400 English phrases were all distinguished by 3 keywords or more except for 4 phrases which differed by 2 keywords. The 15 best keywords in performance order were "the, of, our, yours, in, for, shall, we, you, to, a, have, are, this, no". In French the same set of phrases gave rise to much the same set of results. In this case the 15 best keywords were found to be "de, vous, à, votre, en, notre, par, je, les, pour, nous, nos, des, monsieur, que", and in German they were "wir, sie, mit, die, in, uns, ihre, ihnen, ein, und, für, ihr, nicht, empfehlen, der".

Large phrase separations make the phrase identification system immune to recognition errors. This is illustrated in the following utterance :

"Thank **you** **for** **your** letter last **week**".

The output from the speech recogniser might be :

"**and** **you** **for** * **Mr** * **week**" (* = unrecognised word)

The output phrase differs by 3 keywords from the correct phrase and by 5 keywords from the next closest which is :

"Thank **you** **for** **your** telex asking **for** a quotation".

In this example 3 out of 7 words are misclassified and it was observed that in general, phrases were still correctly identified in spite of error rates as high as 50%. Immunity to such high rates of error are essential if the current generation of commercially available recognition devices are used in the required word spotting mode.

The order of words may also be used to distinguish phrases. To this end pairs of words together with their spacing are extracted according to the same phrase separation criterion, and their presence or otherwise used in the same manner as single keywords. Word pairs are only selected when both keywords are present with different spacings in two or more phrases. This ensures that distinguishing information arising from the word pair is additional to that from the component keywords themselves.

In the example above the separation between the two phrases arises simply from the presence or otherwise of the words "a" and "week". By introducing the word pair "your * * for", where "*" indicates any word, the resulting phrase separation is increased from 2 to 3.

A further improvement in performance is achieved by co-classing frequently confused words. The keywords "to" and "do", for example, are assigned the same label by the recogniser thereby eliminating an important source of errors. Such confusable words are considered to be synonymous during keyword extraction but separate templates are prepared for the recogniser.

The use of continuous speech recognisers would allow the user to speak in a more natural manner. However, such recognisers are currently unable to indicate the number of words spoken and their spacing with any degree of reliability, especially in a word spotting application.

CONCLUSIONS

A speech driven language translation system has been described which operates rapidly and performs remarkably well despite numerous errors from the speech recogniser. At present a phrasebook of typical business communications containing around 400 phrases is being studied. The techniques are however, being extended to other domains of discourse. Higher performance recognisers will allow larger phrasebooks to be implemented, and improved text-to-speech synthesisers will lead to more natural translations which match the voice of the originator.

ACKNOWLEDGEMENT

Acknowledgement is made to the Director of Research of British Telecom for permission to publish this paper.

REFERENCES

1. H Fujisaki, Int. Symp. on Prospects and Problems of Interpreting Telephony, Tokyo, 12th April 1986.
2. P J Hayes et al, Int. J. Man Machine Studies, 19, 231, 1983.
3. D G Bobrow et al, Art. Int., 8, 155, 1977.
4. S J Young, Proc. IEE, 133, Pt E, 305, 1986.
5. R C Schank et al, Am. J. Comp. Ling., 6, 13, 1980.
6. C K Riesbeck et al, Tech. Rep. 78, Comp. Sc. Dept, Yale U., 1976.
7. J Weizenbaum, Comm. ACM, 9, 36, 1966.
8. P J Hayes et al, Am. J. Comp. Ling., 7, 232, 1981.
9. R C Parkison et al, Art. Int., 9, 111, 1977.
10. P J Hayes et al, Proc. Coling '86, 587, 1986.
11. C Schmandt, Proc. Speech Tech '86, 157, New York, April 28-30, 1986.
12. ICAO Annex 10, Vol 2 (Communication Procedures) to the Convention on International Aviation.
13. List of phrases used in the int. telephone service, CCITT, Aug 1965.
14. F W M Stentiford, IEEE Trans. Patt. Anal. Mach. Int., 7, 349, 1985.
15. F W M Stentiford, Int. Conf. Speech I/O, IEE Pub. No. 258, 15, 1986.