

On August 5 British Telecom Research Laboratories announced to the press that they had developed the world's first system for the "instantaneous translation of speech by computer" (no mention was made of the major project now going on in Japan — see story in *LM* November 1985). The prototype provides for the translation of simple sentences (from a limited range of business phrases) into German, Spanish, Swedish and Italian. The reverse capability is being developed.

This will then also make possible translation between any pair of these languages, such as French–German, Swedish–Italian.

Each speaker has a microphone linked to a Merlin 5200 personal microcomputer. These are connected by a telephone circuit capable of handling computer data.

The first participant speaks a sentence in English into his microphone, saying each word clearly and deliberately. The computer repeats the sentence in its own synthetic voice to check that it has understood correctly.

When this is confirmed — by saying the single word 'yes' — the originating computer sends the message to the distant computer which translates and speaks it in, say, French in its own synthetic voice.

When the French speaker replies, the process is repeated in the reverse direction.

The system is based on a set of more than 400 phrases in common business use stored in each computer's memory. Although this involves a vocabulary of more than 1,000 words, the computers are programmed to recognise only 100 key words. These are used to identify the appropriate phrases, reducing the word-recognition task required.

The system also recognises spoken proper names — such as John Smith, for example — and makes no attempt to translate them, e.g. rendering "Mr White" into "Monsieur Blanc". Instead the names are repeated in the original speaker's voice in the translation.

The speech translation project is sponsored at the research laboratories by British Telecom

British Telecom announce voice translation prototype



British Telecom publicity photograph for the new system

International, which is now looking into the possibilities for carrying out a trial with another country.

Three technologies are required to produce a speech translation: (1) a speech recognition device to tell the computer what has been said; (2) a "phrase book" prepared by expert linguists to provide an accurate translation; and (3) a speech synthesiser to enunciate the translation.

British Telecom Research Laboratories have been working on automatic computer translation of text since 1984.

Studies have shown that speech recognition is especially prone to error, but that errors can be avoided if the domain of discourse is restricted. This enables the number of

distinct messages to be kept within the number that can be listed conveniently in a phrase book.

It has also been found that much of the information contained in a message resides in no more than a handful of key words which may be automatically selected for any given phrase book. This means that the speech recogniser may be instructed to recognise only a few truly key words, and in consequence is more likely to get those words right. Several key words have to be incorrectly recognised before phrases become confused.

This provides the system with some immunity not only to recognition errors but also to the variability of the speaker's mode of expression.

A set of more than 400 business letter

phrases in several languages, with total vocabularies of over 12,000 words, has been analysed and key words extracted. All but four of these phrases in English contain at least three of the key words needed to distinguish one phrase from another.

This means that any one of the phrases can be identified reliably by recognising the appropriate combination of three or more key words. All 400 phrases can be covered by just 100 key words.

The booklet of phrases has been prepared by expert linguists and therefore the translation of any phrase can always be guaranteed to be accurate.

The British Telecom Research Laboratories' two-way speech translation system between English and French is based on these concepts:

Speaker-dependent isolated-word technology is at present used so that the input speech is far from fluent. The recogniser must be trained to the voice of each speaker if an acceptable performance is to be obtained.

Phrase parameters such as dates are processed by first identifying the basic phrases using key word recognition. The position of the date in the phrase can then usually be deduced, and a second recognition scan is carried out over a recording of the original utterance, but this time only recognising "date" words. This manoeuvre increases the effective vocabulary of the recogniser without degrading performance.

Names are located in the same manner but no recognition is attempted and the actual speech is transmitted to the other computer. The spoken name is then embedded in the synthesised translation.

In the future speech recognition devices will become speaker-independent and handle continuous speech with larger vocabularies. Recognition of intonation, stress and prosody will reveal additional levels of meaning which are all used in day-to-day speech and may need to be translated. Synthesisers will produce more natural sounding speech which can be matched to the voice of the original speaker.

The new system will be on show at the Telecom 87 international telecommunications exhibition in Geneva this month (October 1987).