DOCUMENT RESUME

ED 413 747 FL 024 768

AUTHOR	de Sopena, Luis
TITLE	Speech Recognition: A General Overview.
PUB DATE	1995-00-00
NOTE	6p.; In: Language Resources for Language Technology: Proceedings of the TELRI (Trans-European Language Resources Infrastructure) European Seminar (1st, Tihany, Hungary, September 15-16, 1995); see FL 024 759.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Computational Linguistics; *Computer Software; *Discourse Analysis; Foreign Countries; Information Technology; *Language Processing; *Language Research; Linguistic Theory; *Oral Language; Trend Analysis
IDENTIFIERS	*Speech Recognition

ABSTRACT

Speech recognition is one of five main areas in the field of speech processing. Difficulties in speech recognition include variability in sound within and across speakers, in channel, in background noise, and of speech production. Speech recognition can be used in a variety of situations: to perform query operations and phone call transfers; for data entry; for command and control operations; and in dictation. Technical characteristics of speech recognition systems depend on several variables, the most important of which are vocabulary size, speaker dependence, speaker mode, domain dependence, and multiple language support. Knowledge sources are based on three models: set of phonemes (acoustic); word lexicon; and language. The objective of the speech recognition process is to determine the sequence of words that most probably caused the observed sequence of acoustic vectors. Currently, speech recognition systems can recognize a large number of words, recognize discrete speech, handle 70-100 words per minute, and handle several languages with a high recognition rate. In the future, speech recognition systems will be able to handle any speaker without need for training, continuous speech, very large vocabularies, telephone communication, and natural language understanding. (MSE)

Speech Recognition
A General Overview

Luis de Sopeña

IBM S. A.
Madrid Scientific Centre
Santa Hortensia 26–28
E-280 002 Madrid
Tel.: +341 397 5752
Fax: +341 519 3990
E-mail: Lsopena@vnet.ibm.com

1. Areas of Speech Processing

There are five main areas in the field of Speech Processing:
1) *Speech Coding* deals with the compression of the digital representation of the speech signal in order to facilitate economical transmission or storage.
2) In *Speech Synthesis*, a synthetic speech signal is created from preexisting text with an attempt at reaching maximum intelligibility and naturalness.
3) Using techniques for *Speaker Identification*, the machine identifies the speaker by his/her voice in order to ensure restricted access to information, computer, or the physical premises.
4) In *Speech Recognition*, the information in a spoken message is identified so as to have the computer perform the corresponding command or transcribe in written form the dictated text.
5) Finally, *Spoken Language Translation* deals with two-way communication via speech: a spoken message is identified, translated into a different language and this translation synthesised in speech form, in order, e.g., to enable a dialogue between speakers of different languages.

2. Difficulties in Speech Recognition

There are some well-known difficulties in the field of speech recognition, shown in the list below:

- The variability of sounds (words, phrases, subword units), within a single speaker and across different speakers.
- The variability of channel, depending on the characteristics of the different types of microphones.
- The variability of background noise: side conversations, street noise, telephone rings, etc.
- The variability of speech production, which adds spurious sounds to words proper (mouth clicks, hesitations, breath noise.)

3. Main Functions of Speech Recognition

Speech recognition can be used in a variety of situations:
1) To perform *Query* operations, such as the consultation via telephone of a bank for account balances, the consultation of phone information lines for theatre schedules and the like, and also for phone call transfers.

2) *Data entry* situations may include the giving of a credit card number, dialing from mobile phones, filling out forms, and booking airline reservations.
3) *Command and Control* operations in which speech recognition is important occur when the hands and/or eyes are busy, during menu navigation and machine control, and while completing dark room work.
4) Speech recognition plays a key role in *dictation* when entering free text into a computer via speech.

4. Technical Characteristics of Speech Recognition Systems

The technical characteristics of speech recognition systems depend on several variables, the most important of which are the following:
1) The *vocabulary size* can range from small (10-100 words) for simple commands, to medium (1000 words) for form filling, or to large (more than 20 000) for such complex situations as dictation.
2) Other than vocabulary size, the *speaker dependence* of a given system can vary from being trained to a specific speaker, to being adaptive to each user as (s)he speaks, or even speaker independent.
3) The *speaking mode* varies between continuous text and isolated words, where pauses between words are needed for an adequate recognition.
4) Speech recognition systems can be *domain dependent*, meaning they can only recognize a constrained syntax (e.g., a list of commands or of questions), or independent, where free text can be dictated.
5) *Multiple language support* is also an important characteristic.

5. Knowledge Sources in Speech Recognition

The knowledge sources in speech recognition are based on three different models:
1) *Set of Phoneme Models:*
Reference to the typical sound of a phoneme, specified by the probability distribution of its spectral and temporal properties.
2) *Word Lexicon:*
Represented as a sequence of the above phonemes (Acoustic Model).
3) *Language Model:*
Statistical model extracted from large corpora of texts.

6. Speech Recognition Process

The objective of the speech recognition process is to determine the sequence of words which caused most probably the observed sequence of acoustic vectors (see figure 1).

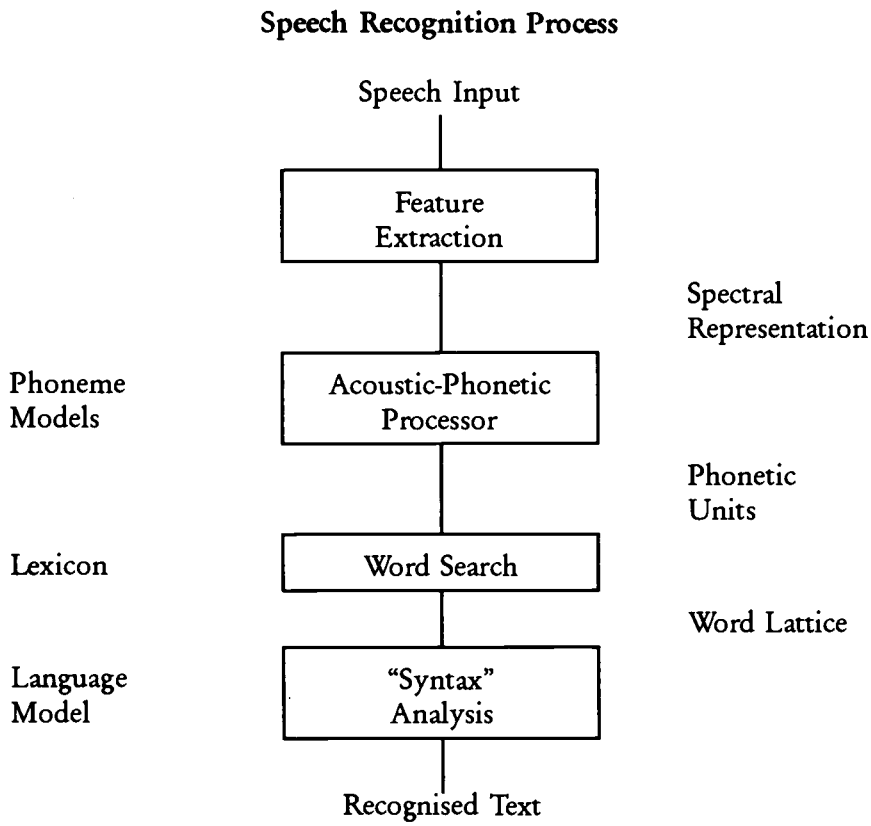

Figure 1. Determine sequence of words which caused MOST PROBABLY the observed sequence of acoustic vectors

7. Speech Recognition Today and Tomorrow

An example of a present-day Speech Recognition system is the *IBM VoiceType Dictation System*. Its most important characteristics are:
- Works on a 486 SX 25
- Recognises more than 30K different words
- Needs a short enrollment process
- Recognises discrete speech (with small pauses between words)
- Able to handle 70–100 words per minute
- Available for 6 languages
- With a very high recognition rate (>96%)

Tomorrow, however, research is promising much more. Speech recognition systems will be able to handle:
- Any speaker, without need for training
- Continuous speech
- Very large vocabularies (more than 250K words)
- With telephone capabilities
- Including natural language understanding
- On Personal Digital Assistants

These systems will be used in dictation, phone mail, DB access, home shopping, translation, and much more.

Most important of all, Speech will be an "enabler", i.e., existing and new applications will be accessible using speech.

8. Main Players in the Field of Speech Recognition

The main players in the field of speech recognition are the following:
- the European Community
- ARPA (Wall Street Journal Contest, Air Travel Info Service (ATIS))
- Industrial Research (IBM in dictation, AT & T for phone services, and many smaller companies)

One of the continual points of discussion in the field of speech recognition is the relative importance of English as compared to other languages. But nonetheless, speech systems are developed for other major languages as well (e.g., French, German, Spanish).

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: TELRI - Proceedings of the First European Seminar:"Language Resources for Language Technology", Tihany, Hungary, Sept. 15 and 16, 1995	
Author(s): Heike Rettig (Ed.)	
Corporate Source:	Publication Date: 1996

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

	The sample sticker shown below will be affixed to all **Level 1** documents	The sample sticker shown below will be affixed to all **Level 2** documents	
[X] Check here **For Level 1 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) *and* paper copy.	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	[ ] Check here **For Level 2 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but *not* in paper copy.
	Level 1	**Level 2**	

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at **Level 1**.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Sign here→ please Signature:	Printed Name/Position/Title: Norbert Volz, M.A. TELRI Project Manager	
Organization/Address: Institut für deutsche Sprache R 5, 6-13 - 68161 Mannheim Postfach 101621 - 68016 Mannheim	Telephone: +49 621 1581-437	FAX: +49 621 1581-415
	E-Mail Address: volz(at)ids-mannheim.de	Date: 28/11/97

(over)