US006085160A

United States Patent [19]

D'hoore et al.

[11] Patent Number: 6,085,160

[45] Date of Patent: Jul. 4, 2000

[54] **LANGUAGE INDEPENDENT SPEECH RECOGNITION**

[75] Inventors: **Bart D'hoore**, Aalter; **Dirk Van Compernolle**, Korbeek-Dijle, both of Belgium

[73] Assignee: **Lernout & Hauspie Speech Products N.V.**, Ieper, Belgium

[21] Appl. No.: **09/113,589**

[22] Filed: **Jul. 10, 1998**

[51] **Int. Cl.**[7] **G10L 5/04**

[52] **U.S. Cl.** **704/256**; 704/2; 704/277

[58] **Field of Search** 704/251, 254, 704/255, 243, 256, 2, 277

[56] **References Cited**

U.S. PATENT DOCUMENTS

5,540,589	7/1996	Waters	704/246
5,717,743	2/1998	McMahan et al.	704/244
5,758,023	5/1998	Bordeaux	704/232
5,768,603	6/1998	Brown et al.	704/232
5,882,202	3/1999	Sameth et al.	704/8
5,915,001	6/1999	Uppaluru	379/88.22
5,963,892	10/1999	Tanka et al.	704/2
5,963,903	10/1999	Hon et al.	704/254

FOREIGN PATENT DOCUMENTS

DE 196 34 138	2/1998	Germany .
WO 98/11534	3/1998	WIPO .

OTHER PUBLICATIONS

Bub, U. et al. "In–Service Adapation of Multilingual Hidden–Markov–Models", *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP '97*), Apr. 21, 1997, pp. 1451–1454.

Constantinescu, A. et al, "On Cross–Language Experiments and Data–Driven Units for ALISP (Automatic Language Independent Speech Processing", *Proceedings IEEE Workshop on Automatic Speech Recognition and Understanding,* Dec. 14–17, 1997, pp. 606–613.

Joachim Kohler, "Multi–Lingual Phoneme Recognition Exploiting Acoustic–Phonetic Similarities of Sounds", pp. 2195–2198.

Weng, et al., "A Study of Multilingual Speech Recognition", ESCA, Eurospeech97, pp. 359–362.

Schultz, et al., "Fast Bootstrapping of LVCSR Systems With Multilingual Phoneme Sets", ESCA, Eurospeech97, pp. 371–374.

Schultz, et al., Japanese LVCSR on the Spontaneous Scheduling Task With Janus–3, ESCA, Eurospeech97, pp. 367–370.

Bonaventura, et al., "Multilingual Speech Recognition for Flexible Vocabularies", ESCA, Eurospeech97, pp. 355–358.

Jayadev Billa, et al., "Multilingual Speech Recognition: The 1996 Byblos Callhome System", ESCA Eurospeech97, pp. 363–366.

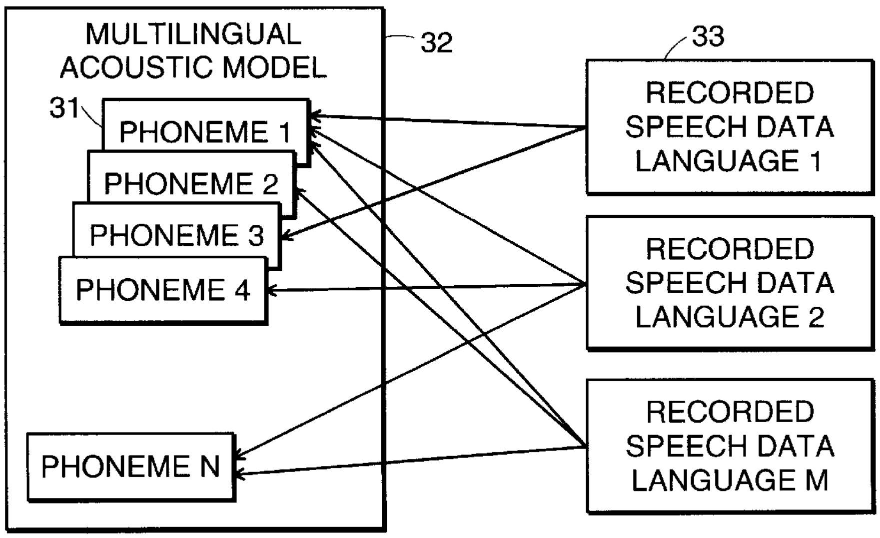
Wang, Chao, et al., "Yinhe: A Mandarin Chinese Version of the Galaxy System", ESCA, Eurospeech97, pp. 351–354.

Primary Examiner—David R. Hudspeth
Assistant Examiner—Susan Wieland
Attorney, Agent, or Firm—Bromberg & Sunstein LLP

[57] **ABSTRACT**

A speech recognition system uses language independent acoustic models derived from speech data from multiple languages to represent speech units which are concatenated into words. In addition, the input speech signal which is compared to the language independent acoustic models may be vector quantized according to a codebook which is derived from speech data from multiple languages.

26 Claims, 3 Drawing Sheets



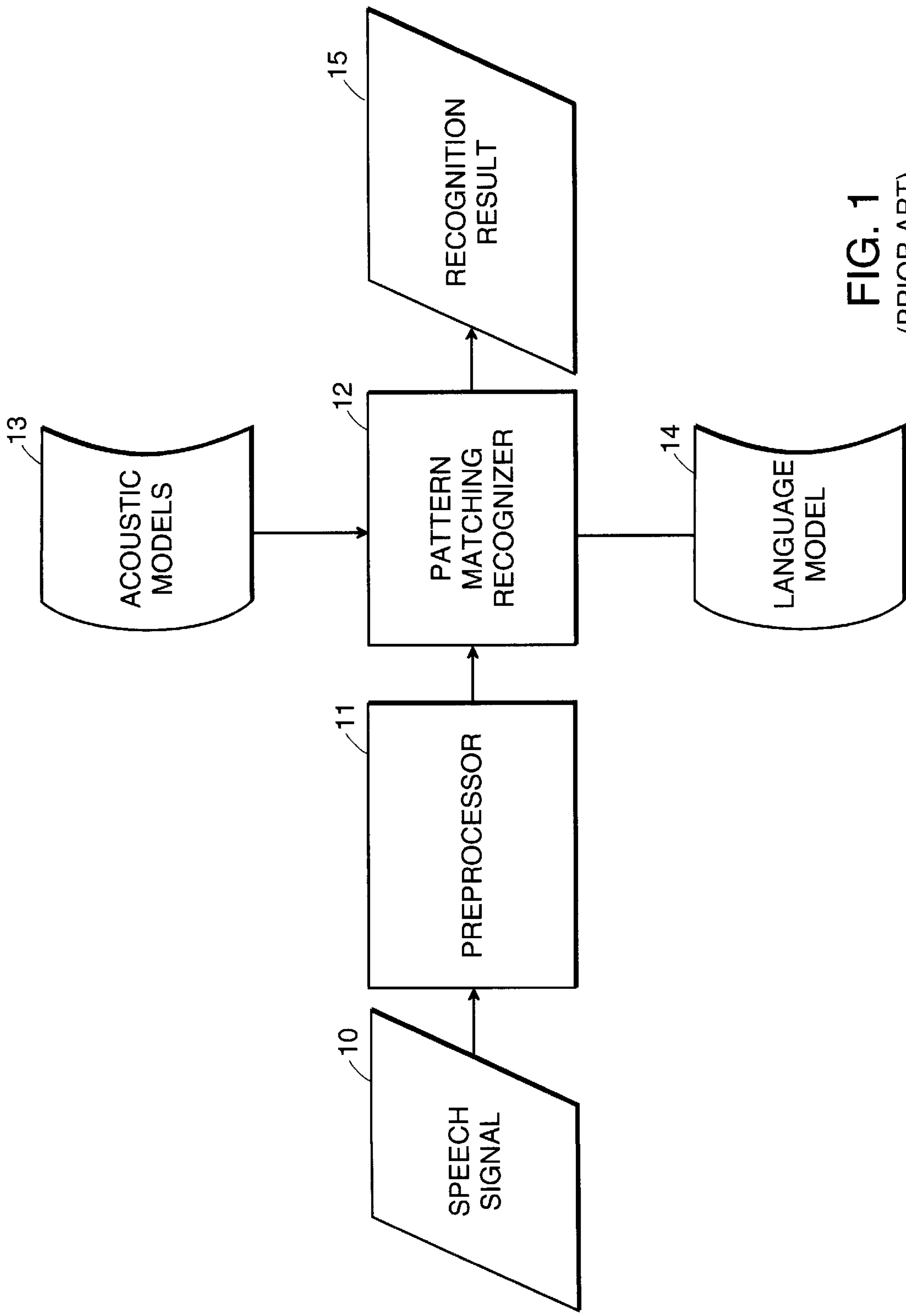


FIG. 1
(PRIOR ART)

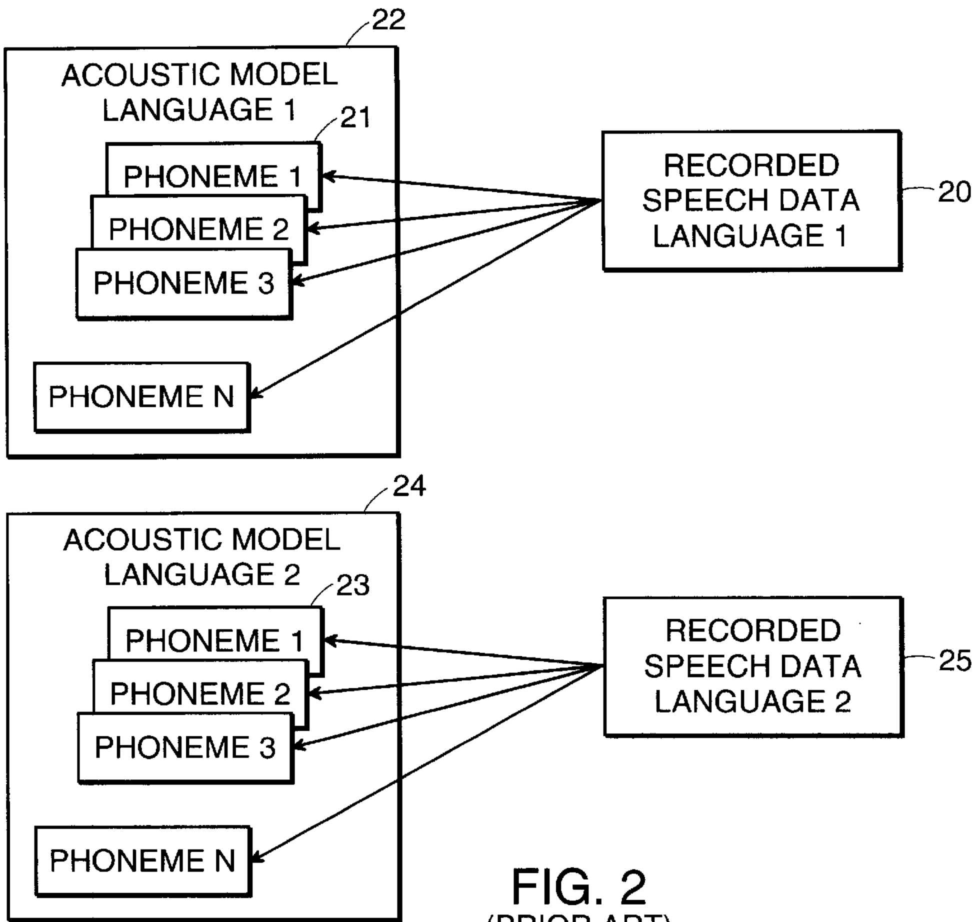


FIG. 2
(PRIOR ART)

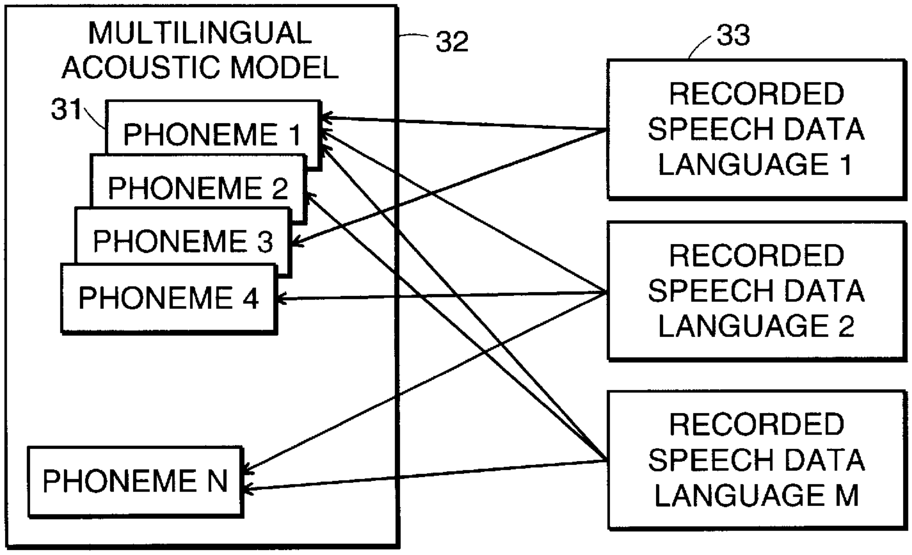


FIG. 3

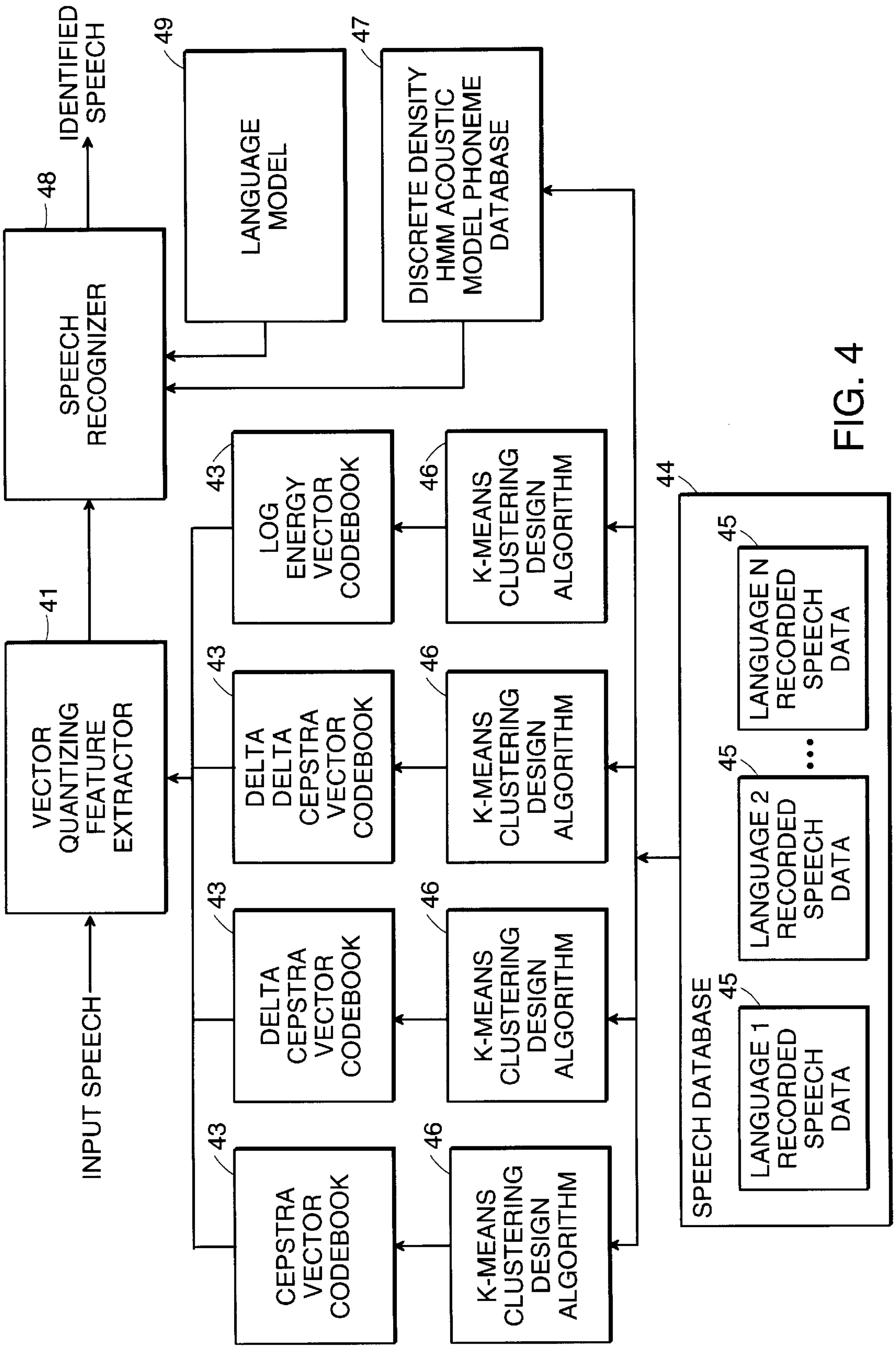

IDENTIFIED SPEECH
48
SPEECH RECOGNIZER
49
LANGUAGE MODEL
47
DISCRETE DENSITY HMM ACOUSTIC MODEL PHONEME DATABASE
41
VECTOR QUANTIZING FEATURE EXTRACTOR
INPUT SPEECH
43
LOG ENERGY VECTOR CODEBOOK
43
DELTA DELTA CEPSTRA VECTOR CODEBOOK
43
DELTA CEPSTRA VECTOR CODEBOOK
43
CEPSTRA VECTOR CODEBOOK
46
K-MEANS CLUSTERING DESIGN ALGORITHM
46
K-MEANS CLUSTERING DESIGN ALGORITHM
46
K-MEANS CLUSTERING DESIGN ALGORITHM
46
K-MEANS CLUSTERING DESIGN ALGORITHM
44
SPEECH DATABASE
45
LANGUAGE 1 RECORDED SPEECH DATA
45
LANGUAGE 2 RECORDED SPEECH DATA
45
LANGUAGE N RECORDED SPEECH DATA

FIG. 4

LANGUAGE INDEPENDENT SPEECH RECOGNITION

TECHNICAL FIELD

The present invention relates to speech recognition systems.

BACKGROUND ART

Current speech recognition systems support only individual languages. If words of another language need to be recognized, acoustic models must be exchanged. For most speech recognition systems, these models are built, or trained, by extracting statistical information from a large body of recorded speech. To provide speech recognition in a given language, one typically defines a set of symbols, known as phonemes, that represent all sounds of that language. Some systems use other subword units more generally known as phoneme-like units to represent the fundamental sounds of a given language. These phoneme-like units include biphones and triphones modeled by Hidden Markov Models (HMMs), and other speech models well known within the art.

A large quantity of spoken samples are typically recorded to permit extraction of an acoustic model for each of the phonemes. Usually, a number of native speakers—i.e., people having the language as their mother tongue—are asked to record a number of utterances. A set of recordings is referred to as a speech database. The recording of such a speech database for every language one wants to support is very costly and time consuming.

SUMMARY OF THE INVENTION

(As used in the following description and claims, and unless context otherwise requires, the term "language independent" in connection with a speech recognition system means a recognition capability that is independently existing in a plurality of languages that are modeled in the speech recognition system.)

In a preferred embodiment of the present invention, there is provided a language independent speech recognition system comprising a speech pre-processor, a database of acoustic models, a language model, and a speech recognizer. The speech pre-processor receives input speech and produces a speech-related signal representative of the input speech. The database of acoustic models represent each subword unit in each of a plurality of languages. The language model characterizes a vocabulary of recognizable words and a set of grammar rules, and the speech recognizer compares the speech-related signal to the acoustic models and the language model, and recognizes the input speech as a specific word sequence of at least one word.

In a further and related embodiment, the speech pre-processor comprises a feature extractor which extracts relevant speech parameters to produce the speech-related signal. The feature extractor may include a codebook created using speech data from the plurality of languages, and use vector quantization such that the speech-related signal is a sequence of feature vectors.

Alternatively, or in addition, an embodiment may create the acoustic models using speech data from the plurality of languages. The subword units may be at least one of phonemes, parts of phonemes, and sequences of phonemes. The vocabulary of recognizable words may contain words in the plurality of languages, including proper nouns, or words in a language not present in the plurality of languages, or foreign-loan words. In addition, the words in the vocabulary of recognizable words may be described by a voice print comprised of a user-trained sequence of acoustic models from the database. Such an embodiment may further include a speaker identifier which uses the voice prints to determine the identity of the speaker of the speech input.

In yet another embodiment, the speech recognizer may compare the relevant speech parameters to acoustic models which represent subword units in a first language in the plurality of languages, and then recognize the speech input as a specific word sequence of at least one word in a second language in the plurality of languages so that input speech from a non-native speaker may be recognized.

Another embodiment of the present invention includes a computer-readable digital storage medium encoded with a computer program for teaching a foreign language to a user which when loaded into a computer operates in conjunction with an embodiment of the language independent speech recognition system described.

Embodiments of the present invention may also include a method of a language independent speech recognition system using one of the systems described above.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be more readily understood by reference to the following detailed description taken with the accompanying drawings, in which:

FIG. **1** illustrates the logical flow associated with a typical speech recognition system.

FIG. **2** illustrates acoustic models of phonemes for multiple languages according to prior art.

FIG. **3** illustrates multi-language acoustic models using a universal set of phonemes according to a preferred embodiment.

FIG. **4** illustrates a speech recognition system according to a preferred embodiment.

DETAILED DESCRIPTION OF SPECIFIC EMBODIMENTS

Operation of a typical speech recognition engine according to the prior art is illustrated in FIG. **1**. A speech signal **10** is directed to a pre-processor **11**, where relevant parameters are extracted from the speech signal **10**. The pattern matching recognizer **12** tries to find the best word sequence recognition result **15** based on acoustic models **13** and a language model **14**. The language model **14** describes words and how they connect to form a sentence. It might be as simple as a list of words in the case of an isolated word recognizer, or as complicated as a statistical language model for large vocabulary continuous speech recognition. The acoustic models **13** establish a link between the speech parameters from the pre-processor **11** and the recognition symbols that need to be recognized. In medium and large vocabulary systems, the recognition symbols are phonemes, or phoneme-like units, that are concatenated to form words. Further information on the design of a speech recognition system is provided, for example, in Rabiner and Juang, *Fundamentals of Speech Recognition* (hereinafter "Rabiner and Juang"), Prentice Hall 1993, which is hereby incorporated herein by reference.

In a prior art system, as illustrated in FIG. **2**, for any given Language **1**, Language **1**-specific recorded speech data **20** is used to generate acoustic models **22** which represent each phoneme **21** in the language. For any other given Language **2**, Language **2**-specific recorded speech data **25** is used to

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.