SPEECH RECOGNITION

VAMSI YALAVARTHI (2015MCS2358)
ARAVIND BHURIYA (2012CS50280)

CONTENTS

- Classes of speech.
- Principle of speech recognition.
- Different approaches to speech recognition.
- Products available commercially.

CLASSES OF SPEECH

- There are four classes:-

1. **Isolated Words:-** It accepts single word or single utterance at a time. This system has “listen/non-listen” states, where they require the speaker to wait between utterances.

2. **Connected Words:-** They allow separate utterances to be run together with minimal pause between them.

3. **Continuous Speech:-** Users speak naturally and recognisers determine the content, but it is difficult to determine the utterance boundaries.

4. **Spontaneous Speech:-** We should be able to handle a variety of natural speech features such as words being run together.

PRINCIPLE OF SPEECH RECOGNITION

- The speech recognition system is essentially a pattern recognition system, including feature extraction and pattern matching.

- It takes an audio stream as input and turns it into a text transcription.

- The speech recognition process can be thought of as having a front end and a back end.

- The **front** end processes the audio stream, isolating segments of sound that are probably speech and converting them into a series of numeric values that characterize the vocal sounds in the signal.

- The **back** end is a specialized search engine that takes the output produced by the front end and searches across three databases.

PRINCIPLE OF SPEECH RECOGNITION

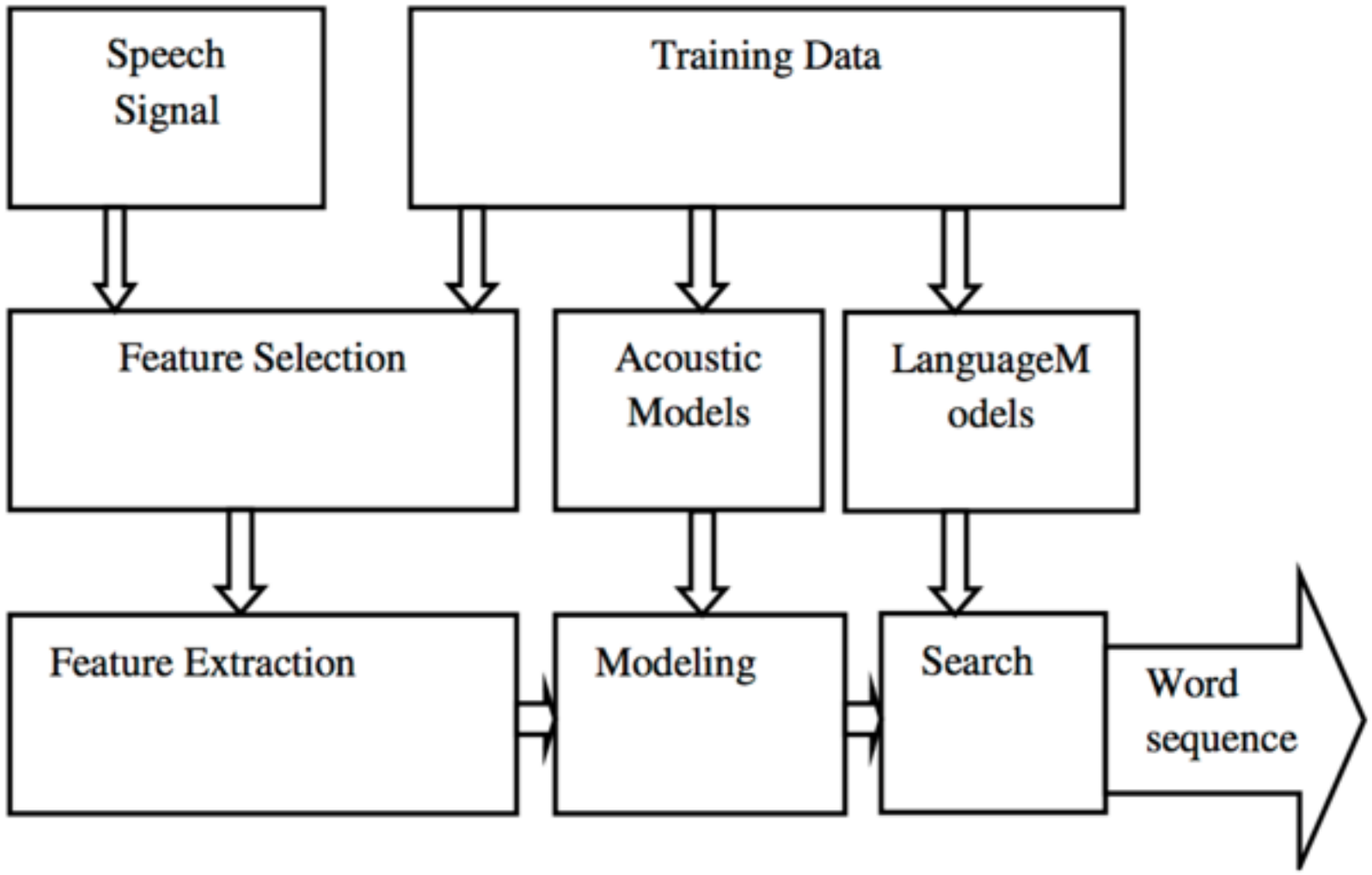


Figure 2: Principle of Speech Recognition

DIFFERENT APPROACHES TO SPEECH RECOGNITION

ACOUSTIC PHONETIC APPROACH:-

- There exist finite, distinctive phonetic units (phonemes) in spoken language.
- These units are broadly characterized by a set of acoustics properties that are manifested in the speech signal over time.
- The first step is the spectral analysis of the speech combined with a feature detection that converts the spectral measurements to a set of features that describe the broad acoustic properties of the different phonetic units.
- The speech signal is segmented into stable acoustic regions, followed by attaching one or more phonetic labels to each segmented region, resulting in a phoneme lattice characterization of the speech.

DIFFERENT APPROACHES TO SPEECH RECOGNITION

PATTERN RECOGNITION APPROACH:-

A speech pattern representation can be in the form of a speech template or a statistical model (e.g a Hidden Markov Model) and can be applied to a sound (smaller than a word), a word, or a phrase.

I. Template based approach

- A collection of prototypical speech patterns are stored as reference patterns which represents the dictionary of candidate words.
- An unknown spoken utterance is matched with each of these reference templates and a category of the best matching pattern is selected.

II. Stochastic approach

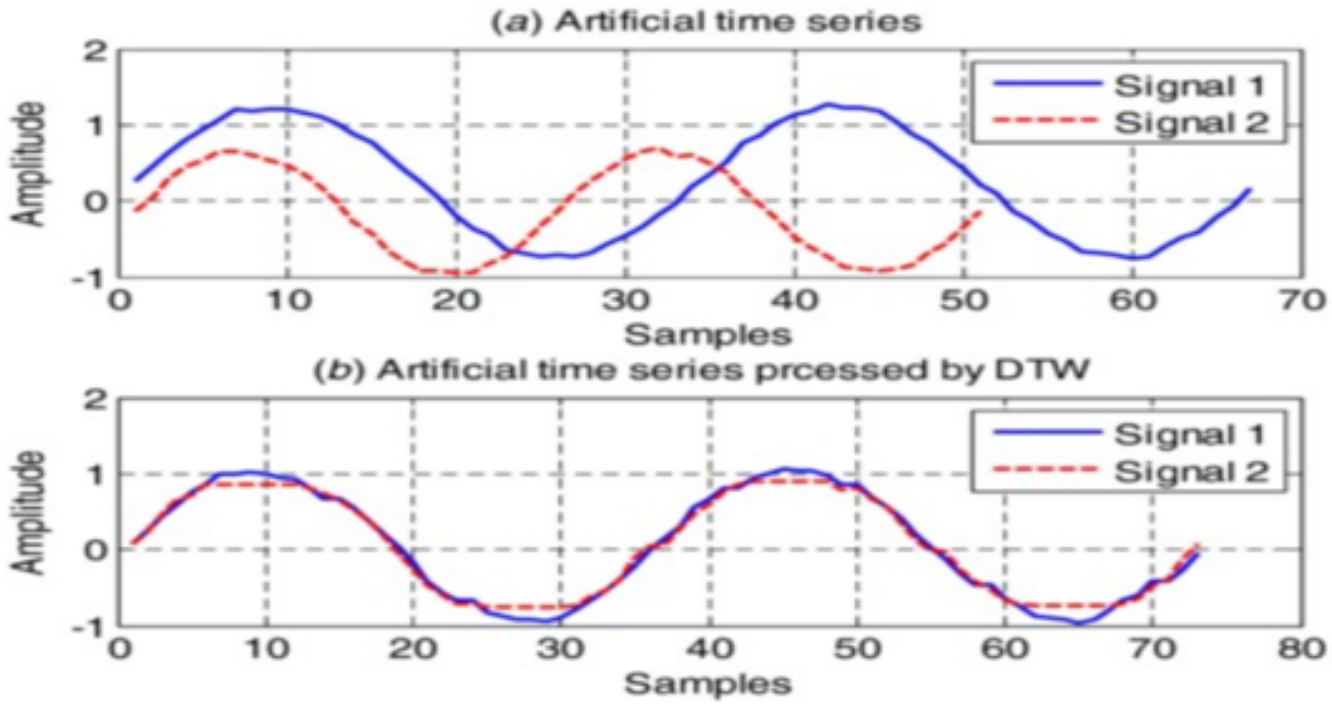
- Stochastic modelling entails the use of probabilistic models to deal with uncertain or incomplete information.
- The most popular stochastic approach today is Hidden Markov modelling.
- A HMM is characterized by a finite state markov model and a set of output distributions.

DIFFERENT APPROACHES TO SPEECH RECOGNITION

DYNAMIC TIME WARPING:-

- It is an algorithm for measuring similarity between two temporal sequences which may vary in time or speed.
- DTW is a method that calculates an optimal match between two given sequences.

- DTW has been applied to temporal sequences of video, audio, and graphics data, indeed any data which can be turned into a linear sequence can be analyzed with DTW.



DIFFERENT APPROACHES TO SPEECH RECOGNITION

ARTIFICIAL INTELLIGENCE APPROACH:-

- Artificial Intelligence approach is a hybrid of the acoustic phonetic approach and pattern recognition approach.

- In its pure form, knowledge engineering design involves the direct and explicit incorporation of expert's speech knowledge into a recognition system.

- This form of knowledge based system enhancement has contributed considerably to the design of all successful strategies reported.

- It plays an important role in the selection of a suitable input representation, the definition of units of speech, or the design of the recognition algorithm itself.

PRODUCTS AVAILABLE COMMERCIALLY

Alexa - Amazon Echo:-

- It is capable of voice interaction, music playback, making to-do lists, setting alarms, streaming podcasts, playing audiobooks, and providing weather, traffic and other real time information.

- It can also control several smart devices using itself as a home automation hub.

- Even when it's alseep, echo is still listening for the trigger phrase that will turn it on and start streaming your voice to the cloud.

Applications:-

- In a few years, Amazon is planning to fill homes with "smart" clocks, refrigerators, TVs and security systems.

- Together, they'll help voice assistants break out of single devices to be available anywhere in houses, cars or workplaces.

PRODUCTS AVAILABLE COMMERCIALLY

Windows Speech Recognition:-

- It is a speech recognition component developed by Microsoft and introduced in the Windows Vista operating system that enables the use of voice commands to perform operations, such as the dictation of text, within applications and the operating system itself.

Applications:-

- With Windows speech recognition, users can dictate text within documents and e-mail messages, fill out forms, control the operating system user interface, perform keyboard shortcuts, and move the mouse cursor.

PRODUCTS AVAILABLE COMMERCIALLY

TrulyNatural:-

- TrulyNatural is the first embedded large vocabulary continuous speech recognizer system which can provide the same state-of-the-art performance as cloud based technologies, but on the device.
- Many applications don’t need, or don’t want to rely on, a cloud based connection to do high performance speech recognition.
- Connections can frequently fail or cause fatal delays, and recent concerns have been expressed about sending personal data to the clouds where it could be stolen or used for purposes undesirable to the owner. TrulyNatural is the solution for these needs.
- It uses a neural network with deep learning to achieve this model. These neural networks employ the most recent breakthroughs in speech feature extraction to produce superior accuracy in real world noise.
- Combining this with a small Finite State Transducer (FST) enables storage and accurate processing of multiple large search domains as required in today’s market.

Application:-

- Currently Voice Dial

PRODUCTS AVAILABLE COMMERCIALLY

SIRI:-

- Siri is a computer program that works as an intelligent personal assistant and knowledge navigator, part of Apple Inc.'s iOS, watchOS, macOS, and tvOS operating systems.
- The feature uses a natural language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of Web services.
- The software, both in its original version and as an iOS feature, adapts to the user's individual language usage and individual searches (preferences) with continuing use, and returns results that are individualized.

Applications:-

- Set reminders, answers questions with web search, current weather and traffic conditions, sports scores, biographies etc.

SIMILAR PRODUCTS:-

- Cortana (In windows)
- S-voice (In Samsungs phones)

THANK YOU