Speech signal processing

SPEECH SIGNAL PROCESSING

ABSTRACT

Speech signal processing is just like as the speech processing in which first the signal is studied and then being procesed ithe form of digital processing. It involves the signals like audio signals, image signals, electrocardiogram signals and control system signals. The speech signal processing is the combination of the speech processing and the signal processing. Speech processing is just the study of the signals like audio, image, etc. and then these signals are being processed in the form of digital representation.

1.) INTRODUCTION

Speech signal processingis the study ofspeechsignalsand the processing methods of these signals. The can be audio, image, control, electrocardiogram signals, etc. The signals are usually processed in adigitalrepresentation, so that the speech processing can be regarded as a special case ofdigital signal processing, applied to speech signal. They are also very close tothe natural language processing(NLP), as its input can come from / output can go to NLP applications. There is an example like text to speech signal which use an information extraction techniques. It refers to the acquisition, manipulation, storage, transfer and output of vocal utterances by a computer. The main applications of speech signal processing are:

1. 1) Speech recognition

1. 2) Speech synthesis

1. 3) Speech compression

The speech signal processing is the combination of the speech processing and the signal processing.

Speech processing is just the study of the signals like audio, image, etc. and then these signals are being processed in the form of digital representation. It is divided into the following five categories: speech coding, speech recognition, voice analysis, speech synthesis and speech enhacement.

Signal processing is an area ofelectrical engineeringandapplied mathematicsthat deals with operations on or analysis of signals, in either discrete or continuous time to perform useful operations on those signals. Depending upon the application, a useful operation could be filtering, spectral analysis, data compression, data transmission, denoising, prediction, smoothing, deblurring, tomographic reconstruction, identification, classification, or a variety of other operations. Signals of interest can includesound, images, time-varying measurement values andsensordata, for example the biological data such as the electrocardiogram signals, the control systemsignals, telecommunicationtransmissionsignals such as radio signals, and many others. Signalsare analog or digital electrical representations of time-varying or spatial-varying physical quantities. In the context of signal processing, arbitrary binary data streams and on-off signals are not considered as signals, but only analog and digital signals that are representations of analog physical quantities. The signal processing is

categorised into three types which are: audio signal processing, discrete time processing and the digital signal processing.

1. 1) SPEECH RECOGNITION

It is also called voice recognition which focuses on capturing the human voice as a digital sond wave and converting it into the format which can be read by computer. It is also known as automatic speech recognition or the computer speech rscognition. It converts the voice of human being into the machine readable input like computers. The term " voice recognition" is sometimes used to refer to speech recognition where the recognition system is trained to a particular speaker - as is the case for most desktop recognition software, hence there is an aspect ofspeaker recognition, which attempts to identify the person speaking, to better recognise what is being said. Speech recognition is a broad term which means it can recognise almost anybodys speech - such as a callcentre system designed to recognise many voices. Voice recognition is a system trained to a particular user, where it recognises their speech based on their unique vocal sound.

The applications of speech recognition are as follows:

a.) Health care:

In the health care, voice recognition technologies are sidely used. Speech recognition can be implemented in front-end or back-end of the medical documentation process. Front-End SR is where the provider dictates into a speech-recognition engine, the recognized words are displayed right after they are spoken, and the dictator is responsible for editing and signing off on the document. It never goes through an MT/editor. Back-End SR or Deferred SR is where the provider dictates into a digital dictation system, and the

voice is routed through a speech-recognition machine and the recognized draft document is routed along with the original voice file to the MT/editor, who edits the draft and finalizes the report. Deferred SR is being widely used in the industry currently. ManyElectronic Medical Records(EMR) applications can be more effective and may be performed more easily when deployed in conjunction with a speech-recognition engine. Searches, queries, and form filling may all be faster to perform by voice than by using a keyboard.

b.) Military:

Substantial efforts have been devoted in the last decade to the test and evaluation of speech recognition in fighter aircraft. Of particular note are the U. S. program in speech recognition for the Advanced Fighter Technology Integration (AFTI)/F-16aircraft (F-16 VISTA), the program in France on installing speech recognition systems onMirageaircraft, and programs in the UK dealing with a variety of aircraft platforms. In these programs, speech recognizers have been operated successfully in fighter aircraft with applications including: setting radio frequencies, commanding an autopilot system, setting steer-point coordinates and weapons release parameters, and controlling flight displays. Generally, only very limited, constrained vocabularies have been used successfully, and a major effort has been devoted to integration of the speech recognizer with the avionics system.

Some important conclusions from the work were as follows:

* Speech recognition has definite potential for reducing pilot workload, but this potential was not realized consistently.

* Achievement of very high recognition accuracy (95% or more) was the most critical factor for making the speech recognition system useful— with lower recognition rates, pilots would not use the system.

* More natural vocabulary and grammar, and shorter training times would be useful, but only if very high recognition rates could be maintained.

Laboratory research in robust speech recognition for military environments has produced promising results which, if extendable to the cockpit, should improve the utility of speech recognition in high-performance aircraft.

Working with Swedish pilots flying in theJAS-39Gripen cockpit, Englund (2004) found recognition deteriorated with increasing G-loads. It was also concluded that adaptation greatly improved the results in all cases and introducing models for breathing was shown to improve recognition scores significantly. Contrary to what might be expected, no effects of the broken English of the speakers were found. It was evident that spontaneous speech caused problems for the recognizer, as could be expected. A restricted vocabulary, and above all, a proper syntax, could thus be expected to improve recognition accuracy substantially.

TheEurofighter Typhooncurrently in service with the UKRAFemploys a speaker-dependent system, i. e. it requires each pilot to create a template. The system is not used for any safety critical or weapon critical tasks, such as weapon release or lowering of the undercarriage, but is used for a wide range of othercockpitfunctions. Voice commands are confirmed by visual and/or aural feedback. The system is seen as a major design feature in the reduction of pilotworkload, and even allows the pilot to assign targets to

himself with two simple voice commands or to any of his wingmen with only five commands.

1. 2) SPEECH SYNTHESIS

The artificial production of human speech is called the speech synthesis. A computer system used for this purpose is called aspeech synthesizer, and can be implemented insoftwareorhardware. It is the reverse process of speech recognition and advances in the area to improve the computer's usability for the visually impaired. Atext-to-speech (TTS)system converts normal language text into speech, other systems rendersymbolic linguistic representationslikephonetic transcriptionsinto speech.

Synthesized speech can be created by concatenating pieces of recorded speech that are stored in adatabase. Systems differ in the size of the stored speech units; a system that storesphonesordiphonesprovides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Alternatively, a synthesizer can incorporate a model of thevocal tractand other human voice characteristics to create a completely " synthetic" voice output.

The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood. An intelligible text-to-speech program allows people withvisual impairmentsorreading disabilitiesto listen to written works on a home computer. Many computer operating systems have included speech synthesizers since the early 1980s.

A text to speech system (TTS) is explained below:

Overview of a typical TTS system

A text-to-speech system is composed of two parts: afront-endand aback-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written-out words. This process is often calledtext normalization. The front-end then assignsphonetic transcriptionsto each word, and divides and marks the text intoprosodic units, likephrases, clauses, andsentences. The process of assigning phonetic transcriptions to words is calledtext-to-speech. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end. The back-end often referred to as thesynthesizer then converts the symbolic linguistic representation into sound.

The application of speech synthesis are:

a.) Accessibility:

Speech synthesis have a technology tool and its application is widely spread in some areas. It allows environmental barriers to be removed for people with a wide range of disabilities. The longest application has been in the use ofscreenreadersfor people withvisual impairment, but text-to-speech systems are now commonly used by people withdyslexiaand other reading difficulties as well as by pre-literate youngsters. They are also frequently employed to aid those with severespeech impairmentusually through a dedicatedvoice output communication aid.

b.) Entertainment:

Speech synthesis techniques are also widely used as entertainment such as games. In 2007, Animo Limited announced the development of a software

application package based on its speech synthesis software FineSpeech, explicitly geared towards customers in the entertainment industries, able to generate narration and lines of dialogue according to user specifications. The application reached maturity in 2008, when NECBiglobeannounced a web service that allows users to create phrases from the voices ofCode Geass: Lelouch of the Rebellion R2characters. Software such asVocaloidcan generate singing voices via lyrics and melody. This is also the aim of the Singing Computer project (which uses theGPLsoftwareLilypondandFestival) to help blind people check their lyric input.

1. 3) SPEECH COMPRESSION

It is very important in the telecommunications area for increasing the amount of information which can be transferred, stored, or heard, for a given set of time and space constraints. The compression of speech signals has many practical applications. One example is in digital cellular technology where many users share the same frequency bandwidth. Compression allows more users to share the system than otherwise possible. Another example is in digital voice storage (e. g. answering machines). For a given memory size, compression allows longer messages to be stored than otherwise.

In the history, the digital speech signals are sampled at a rate of 8000 samples/sec. Each of the sample is represented by 8 bits (using mu-law). This corresponds to an uncompressed rate of 64 kbps (kbits/sec). With current compression techniques (all of which are lossy), it is possible to reduce the rate to 8 kbps with almost no perceptible loss in quality. Further compression is possible at a cost of lower quality. All of the current low-rate

speech coders are based on the principle oflinear predictive coding (LPC)which is presented in the following sections.

Speech compression may also means the two different things i. e. speech coding and time compressed speech.

Now, there is an example that how we speak and how speech comes out from our mouth.

Physical Model of Speech Production

When we speak:

Air is pushed from your lung through your vocal tract and out of your mouth comes speech.

a. For certainvoicedsound, your vocal cords vibrate (open and close). The rate at which the vocal cords vibrate determines thepitchof your voice. Women and young children tend to have high pitch (fast vibration) while adult males tend to have low pitch (slow vibration).

b. For certainfricatives and plosive (or unvoiced)sound, your vocal cords do not vibrate but remain constantly opened.

c. The shape of your vocal tract determines the sound that you make.

d. As you speak, your vocal tract changes its shape producing different sound.

e. The shape of the vocal tract changes relatively slowly (on the scale of 10 msec to 100 msec).

f. The amount of air coming from your lung determines the loudness of your voice.

CONCLUSION

Speech signal processingis the study ofspeechsignalsand the processing methods of these signals. The can be audio, image, control, electrocardiogram signals, etc. The signals are usually processed in adigitalrepresentation. The speech signal processing is the combination of the speech processing and the signal processing. The main applications of speech signal processing are: Speech recognition, Speech synthesis and Speech compression. The artificial production of human speech is called the speech synthesis. Speech processing is just the study of the signals like audio, image, etc. and then these signals are being processed in the form of digital representation.

REFERNCES

[1] http://en. wikipedia. org/wiki/Speech_signal_processing

[2] http://en. wikipedia. org/wiki/Speech_recognition

[3] http://en. wikipedia. org/wiki/Speech_synthesis

[4] http://en. wikipedia. org/wiki/Speech_compression

[5] http://en. wikipedia. org/wiki/Signal_processing

[6] http://en. wikipedia. org/wiki/Speech_processing

[7] http://www. dspecialists. com/en/produkte/audiosysteme/software/sprachsignalverarbeitung. html