ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 10, October 2014

“Speech Translation System for Vernacular Languages”

Ms. A. H. Utgikar,[1] Mr. A. S. Deshpande[2], Mr. K. S. Ambulgekar[3], Mr. K. R. Joshi[4]

Asst. Prof., E & TC Dept., Maharashtra Institute of Technology, Aurangabad, Maharashtra, India[1]

E & TC Dept., Maharashtra Institute of Technology, Aurangabad, Maharashtra, India[234]

ABSTRACT: This paper explains the nascent stage of developing a personalized interpreter. We propose to develop a prototype which uses a speech processing hardware and on translators to provide the user with real time translation. Speech processing hardware works on the principle of ‘compare and forward’, i.e., a database is already stored in the unit which is used for comparing with the input speech and the result is forwarded for further processing. The need arises from the inability of dictionaries and human translators to suit our needs for better communication. In this situation the prototype proposed will suffice the purpose reasonably well and minimize the communication inefficiencies.

Keywords: Speech Recognition HM2007, Language Translator, Microcontroller, Speech Synthesizer APR6016.

I. INTRODUCTION

The global, borderless economy has made it critically important for speakers of different languages to be able to communicate. Speech translation technology – being able to speak and have one’s words translated automatically into the other person’s language – has long been a dream of humankind. Speech translation has been selected as one of the ten technologies that will change the world. There are especially high hopes in Japan for a speech-translation system that can automatically translate one’s everyday speech as use of Japanese has become increasingly international, so such speech-translation technology would be a great boon to the nation.

Automatic speech translation technology consists of three separate technologies: technology to recognize speech (speech recognition); technology to translate the recognized words (language translation); and technology to synthesize speech in the other person’s language (speech synthesis). Recent technological advances have made automatic translation of conversational spoken Japanese, English, and Chinese for travellers practical, and consecutive translation of short, simple conversational sentences spoken one at a time has become possible.

This report starts by affirming the significance of speech-translation technology, and providing an overview of the state of research and development to date, and the history of automatic translation technology. It goes on to describe the architecture and current performance of speech translation systems.

II. SURVEY

Following are the systems implemented for Speech to Text & Speech to Speech Translation commercially available in the market.

A. Text To Speech Translator:

The SP0-512 Text to Speech IC is a pre-programmed microcontroller that accepts English text from a serial connection converts that text to phoneme codes then generates audio. It is ideal for adding a robot voice to your embedded designs.

B. VOICE ACTIVATED PHRASE LOOKUP (Text to Speech System):

Voice activated phrase lookup systems are not true speech translation systems by definition. A typical voice activated phrase lookup system is the Phraselator system. The Phraselator is a one-way device that can recognize a set of pre-defined phrases and play a recorded translation. This device can be ported easily to new languages, requiring only a hand translation of the phrases and a set of recorded sentences. However, such a system severely limits communication as the translation is one way, thus reducing one party’s responses to simple pointing and perhaps yes and no.

C. *SIGMO (Speech to Speech System):*

SIGMO allows real-time translating of 25 languages. It has two modes of voice translation. Set the native language, then the language to translate to. By pressing the first button and speaking the Set phrase SIGMO in turn will instantly translate and pronounce it in a selected language. By pressing the second button, it will translate speech from the foreign language, then instantly speak selected native language.

D. MASTOR (IBM)

MASTOR (Multilingual Automatic Speech-To-Speech Translator) is IBM’s highly trainable speech-to-speech translation system, targeting conversational spoken language translation between English and Mandarin Chinese for limited domains. Depicts the architecture of MASTOR. The speech input is processed and decoded by a large-vocabulary speech recognition system. Then the transcribed text is analysed by a statistical parser for semantic and syntactic features. A sentence-level natural language generator based on maximum entropy (ME) modelling is used to generate sentences in the target language from the parser output. The produced sentence in target language is synthesized into speech by a high quality text-to-speech system.

F. Matrix (ATR)

The Spoken Language Translation Research Laboratories of the Advanced Telecommunications Research Institute International (ATR) has five departments. Each department focuses on a certain area of Speech Translation.

This system can recognize natural Japanese utterances such as those used in daily life, translate them into English and output synthesized speech. This system is running on a workstation or a high end PC and achieved nearly real-time processing. Unlike its predecessor ASURA, ATR-MATRIX is designed for spontaneous speech input, and it is much faster. The current implementation deals with a hotel room reservation task. ATR-MATRIX adopted a cooperative integrated language translation model. Because of its small, light size, and available attachments it is portable and easy to use.

III. PROPOSED SYSTEM

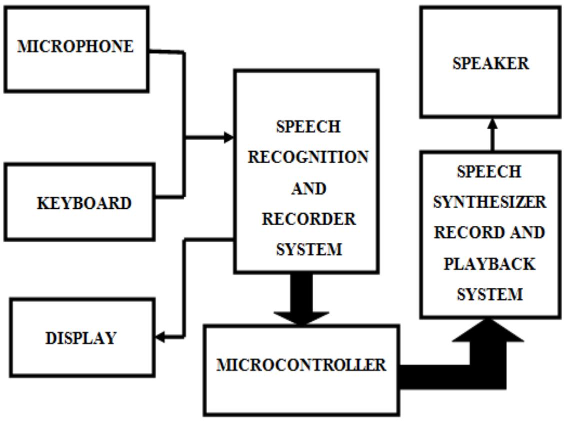
MICROPHONE
KEYBOARD
DISPLAY
SPEECH RECOGNITION AND RECORDER SYSTEM
MICROCONTROLLER
SPEAKER
SPEECH SYNTHESIZER RECORD AND PLAYBACK SYSTEM

Fig. 1. Block Diagram of the system

The figure explains the Block Diagram of Voice to Voice Language Translation System in which the input speech is given through the microphone which then goes to the speech processing unit. This unit processes the input and the word which was spoken is recognized. The input speech first goes to the speech IC of the speech processing unit.

A. Speech Recognition System

Speech Recognition is the process of converting an acoustic signal, captured by microphone or telephone to a set of words. In this system, HM2007 is used as a Speech Recognition unit. The HM2007 is a CMOS voice recognition LSI (Large Scale Integration) circuit. The chip contains an analog front end, voice analysis, regulation, and system control functions. The chip may be used in a stand alone or CPU connected. Some features of HM2007 are as follows:

- Single chip voice recognition CMOS LSI
- Speaker dependent
- External RAM support
- Maximum 40 word recognition (.96 second)
- Maximum word length 1.92 seconds (20 words)
- Microphone support
- Manual and CPU modes available
- Response time less than 300 milliseconds
- 5V power supply

The speech recognition system is a completely assembled and easy to use programmable Speech recognition circuit. Programmable, in the sense that we train the words (or vocal Utterances) we want the circuit to recognize. This board allows you to experiment with many facets of speech recognition technology. It has 8 bit data out which can be interfaced with any Microcontroller for further development.

The input to the system is feed by microphone to speech recognizer circuitry which is used to recognize the words that are already stored in the system. The speech recognition & recording system requires an external memory which is sufficed by a SRAM. Speech recognition & recording system along with the static RAM forms the fundamental block of the speech processing unit. The database is stored in the SRAM and the Speech processing unit is used in the recognition mode where comparison of the input and the database takes place and a particular eight bit BCD address is given as the result. This BCD address is feed to digital data processing unit. The microcontroller used in this system will convert the input address from HM2007 and process it in such a way that the address generated at the output will specify the address of the same word but in the different language, which will be then feed to the APR6016 in order to retrieve the word stored in the synthesizer system.

B. Speech Synthesizer

In this system we are using APR6016 as audio playback and recorder part which is at the output of the system as shown in the block diagram. The APR6016 offers non-volatile storage of voice and/or data in advanced Multi-Level Flash memory. Up to 16 minutes of audio recording and playback can be accommodated. The APR6016 memory array is organized to allow the greatest flexibility in message management and digital storage. The smallest addressable memory unit is called a “sector”. The APR6016 contains 1280 sectors. Sectors 0 through 1279 can be used for analog storage. During audio recording one memory cell is used per sample clock cycle.

The APR 6016 stores voice signals by sampling incoming voice data and storing the sampled signals directly into FLASH memory cells. Each FLASH cell can support voltage ranges from 1 to 256 levels. These 256 discrete voltage levels are the equivalent of eight (28=256) bit binary encoded values. During playback the stored signals are retrieved from memory, smoothed to form a continuous signal and finally amplified before being fed to an external speaker amplifier. Device control is accomplished through an industry standard SPI interface that allows a microcontroller to manage message recording and playback.

The APR 6016 is equipped with an internal squelch feature. The Squelch circuit automatically attenuates the output signal by 6 dB during quiet passages in the playback material. Muting the output signal during quiet passages helps eliminate background noise. Background noise may enter the system in a number of ways including: present in the original signal, natural noise present in some power amplifier designs, or induced through a poorly filtered power supply.

The APR contains a 20 bit op-code register, out off which 14 bits are for the sector address and remaining 5 bits are for the op-code of various instruction. The instructions and there op-code with the summary of the instruction is listed in the table given below:

TABLE I
OPERATIONAL CODES

INSTRUCTION NAME	OP-CODE [OP4-OP0]	SUMMARY
NOP	00000	No Operation
SID	00001	Causes the Silicon ID to be read
STOP	00110	Stop the current Operation
SET_REC	01000	Start a Record Operation from the Sector Address specified
REC	01001	Start a Record Operation from the Current Sector Address specified
SET_PLAY	01100	Start a Playback Operation from the Sector Address specified
PLAY	01101	Start a Playback Operation from the current Sector specified

The audio signal containing the content we wish to record should be fed into the differential inputs ANAIN-, and ANAIN+. After pre-amplification the signal is routed into the anti-aliasing filter.The anti-aliasing filter automatically adapts its response based on the sample rate being used. No external anti-aliasing filter is therefore required. After passing through the anti-alias filter, the signal is fed into the sample and hold circuit which works in conjunction with the Analog Write Circuit to store each analog sample in a flash memory cell. The audio signal containing the content you wish to record should be fed into the differential inputs ANAIN-, and ANAIN+. After pre-amplification the signal is routed into the anti-aliasing filter. The anti-aliasing filter automatically adapts its response based on the sample rate being used. No external anti-aliasing filter is therefore required. After passing through the anti-alias filter, the signal is fed into the sample and hold circuit which works in conjunction with the Analog Write Circuit to store each analog sample in a flash memory cell.

When a SET_REC or REC command is issued the device will begin sampling and storing the data present on ANAIN+ and ANAIN- to the specified sector. After half the sector is used the SAC pin will drop low to indicate that a new command can be accepted. The typical recording sequence is as shown below:

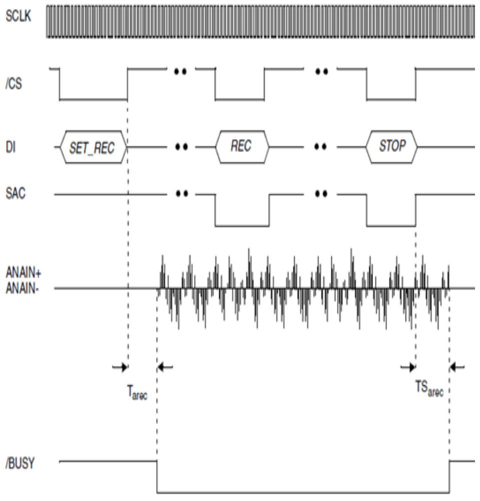

Fig. 2. Typical Recording Sequence

When a SET_PLAY or PLAY command is issued the device will begin sampling the data in the specified sector and produce a resultant output on the AUDOUT, ANAOUT-, and ANAOUT+ pins. After half the sector is used the SAC pin will drop low to indicate that a new command can be accepted. The device will accept commands as long as the SAC pin remains low. Any command received after the SAC returns high will be queued up and executed during the next SAC cycle. Figure 3 shows typical playback sequence:

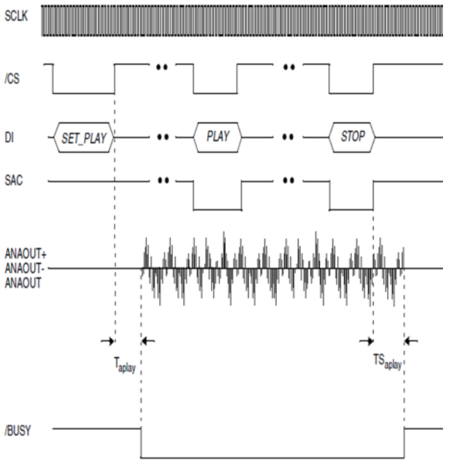

Fig. 3. Typical Playback Sequence

IV. FLOW CHART

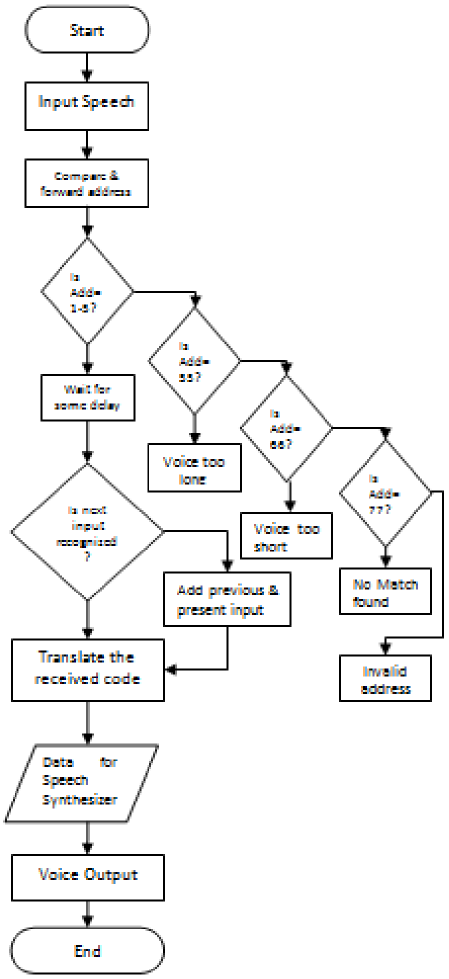

Fig. 4. System flow chart

V. CONCLUSION

Voice to Voice Language Translation system is a device that is designed to bridge the language gap between individuals and foreigners when traveling in our country. The need arises from the inability of dictionaries and human translators to suit our needs for better communication. At present we need 'Personalized Interpreters' which will reduce our dependence on dictionaries and human interpreters. This will reduce the hindrance posed by the language barrier. In this situation the system proposed will suffice the purpose reasonably well and minimize the communication inefficiencies. The system can overcome the real time difficulties of illiterate people and improve their lifestyle.

REFERENCES

[1] 'Sign Language to Speech Translation System Using PIC Microcontroller', Gunasekaran. K1, Manikandan. R2, Senior Assistant Professor2, School of Computing, SASTRA University, Tirumalaisamudram, Tamilnadu, India-613401. guna1kt@gmail.com1, manikandan75@core.sastra.edu2. Volume 5, No 2 Apr-May 2013 [ISSN NO: 0975-4024]

[2] 'Speech to Speech Language Translator', Umeaz Kheradia, Abha Kondwilkar, B.E (Electronics & Telecommunication), Rajiv Gandhi Institute of Technology.umeaz_kheradia17@yahoo.com, bhassk@yahoo.co.in. Volume 2, Issue 12, December 2012 [ISSN 2250-3153]

[3] "Process Speech Recognition System using Artificial Intelligence Technique", Anupam Choudhary, Ravi Kshirsagar, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-5, November 2012.

[4] "An Implementation of Text Dependent Speaker Independent Isolated Word Speech Recognition Using HMM" INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY (IJESRT), Ms. Rupali S Chavan, Dr. Ganesh S. Sable, [September, 2013] ISSN: 2277-9655 Impact Factor: 1.852

[5] Patent US103508 B3 – Voice Activated Language Translator.

[6] Patent Brevetto US6085160 – Language Independent Speech Recognition.

BIOGRAPHY

Ms. A. H. Utgikar M.Tech (Electronics Design Technology) Currently working as Asst. Prof. in Electronics & Telecommunication Engineering Department at G. S. Mandal's Maharashtra Institute of Technology Aurangabad since 2011. Has a experience as a visiting faculty for DESD CDAC and 1.4 years in Industry. Has published a paper on 'Drying of Grapes using Infrared Heating Mechanism' in International Journal of Innovations in Engineering & Technology in August 2013. Also presented a paper on 'Performance analysis of batteries & inverters with polymer as an electrolyte' at National Level Technical paper presentation held at SNJCOE, Nashik. M.Tech Project – 'Design & Development of IR Based Grape Drying System' was Exhibited in International Electronics Expo 2011 by Electronics For You held at Pragati Maidan, New Delhi. Areas of Interest: Speech Processing, Agri Instrumentation & Control, Signal Processing.

Mr. Akshay S. Deshpande U.G. Student Department of Electronics & Telecommunication Engineering G. S. Mandal's Maharashtra Institute of Technology, Aurangabad. Areas of Interest: Speech Processing, Neural Networks

Mr. Keshav S. Ambulgekar U.G. Student Department of Electronics & Telecommunication Engineering G. S. Mandal's Maharashtra Institute of Technology, Aurangabad. Areas of Interest: Speech Processing, Image Processing

Mr. Kedar R. Joshi U.G. Student Department of Electronics & Telecommunication Engineering G. S. Mandal's Maharashtra Institute of Technology, Aurangabad. Areas of Interest: Speech Processing, PCB design, VLSI programing.