Cell Phone For Blind,Deaf And Dumb

J. Nancy Priyadarshini, STUDENT,ME(VLSI DESIGN)

Dep of electronics and communication ,PSNA CET,dindigul,Tamilnadu,INDIA

S. Sivaranjani, STUDENT,ME(VLSI DESIGN)

Dep of electronics and communication ,PSNA CET,dindigul,Tamilnadu, INDIA

B. Vincy Vimala Rani, STUDENT,ME(VLSI DESIGN)

Dep of electronics and communication ,PSNA CET,dindigul-624002,Tamilnadu, INDIA

ABSRACT

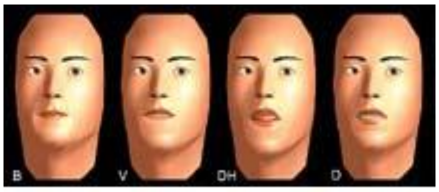
The growth of wireless networks in the past few years can be attributed to rising demand for wireless services such as data, voice, video, and the development of new wireless standards. The rapid wireless technological change has made cell phones reach every corner of the society. The cell phone and wireless technology could be in everyway extended to support an all together deferent echelon of users. The hearing, speech and vision deprived user could be helped to use services of the cell phone as well.This paper explores the possibility of integrating the applications of ***Speech to Animation(STA),Voice recognition*** *to help the blind, deaf and dumb users. Speech to animation technology converts human speech to animated film in real time. The core of this technology is a real time speech recognition engine which converts human voice to phonemes. So ideally, this technology involves generating the animated images based on the phonemes of the received speech content in real time.*

***KEYWORDS*:**

Speech to animation technology, voice recognition, phonemes, speech recognition engine.

INTRODUCTION

As evident, the cell phone has graduated from a device providing the basics of services to accommodating a multitude of applications that can be added according to the needs of the people. A typical cell phone for the deaf and dumb user shall have the STA application in it which can play the animated face in the screen of the phone. It shall also have Video to Speech application and a camera to record the facial (especially lips) movements of the user. An alerter similar to a vibrator can be used to indicate the user of a new call. The blind user needs voice recognition software to take voice commands from the blind user to make a call to a destination.Speech is produced by the movements of specific organs or articulators of the vocal tract. Each distinct speech sound is related to a characteristic positioning of the articulators, and some of their movements are wholly or partially visible on the speaker's face, especially in the region around the mouth, which comprises the upper end of the vocal tract. The simulation of such visible movements during speech production is the key to the generation of realistic speech synchronized 3D facial animation.

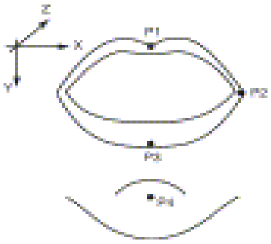


The mouth forms characteristic shapes while uttering certain speech sounds. Here, "Baldy" begins to mouth the sounds "bah," "vah," "thah" (as in the word the), and "dah."

Phonemes are the building blocks of human speech. These are the smallest sound units that make up the words that are spoken. When the sound of a phoneme changes, the change in the word is effected as well. The term viseme, a shorthand for visual phoneme, constitutes a central concept of speech synchronized facial animation. Visemes can be understood as the recognizable visual motor patterns usually common to two or more phonemes. These phonemes are the smallest speech sound units in a language that are capable of conveying a distinction in meaning, such as the ‘p’ in ‘pie’ in contrast with the ‘b’ in ‘bye’. Visemes can be distinguished based on the measurements of movements displayed on the face during speech production to derive geometric models for visemes.

SPEECH TO ANIMATION

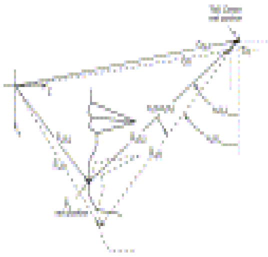
Speech to animation technology converts human speech to animated film in real time.Here a camera captures the lip movements of the blind user and his voice signals are converted to phonemes. These phonemes are converted to visemes and an animated face opens up at the receiving end. The deaf and dumb user at the receiving end can interpret the information by lip reading from the animated face. Audio is probably the most natural modality to recognize speech content and a valuable source to identify a speaker [1]. Video also contains important biometric information, which includes face/lip texture and lip motion information that is correlated with the audio.State-of-art speech recognition systems have been jointly using lip information with audio. For speech recognition, it is usually sufficient to extract the principal components of the lip information and to match the mouth openings–closings with the phonemes of speech.The fusion of audio modality with best lip movements will present good speech recognition engines in real time.



To capture the geometry information during speech production, fiduciary points were marked with white paint on the face of the speaker as shown in the fig. It represents the location and labelling of the four fiduciary points, P1, P2, P3 and P4, for which 3D trajectories were measured and analyzed. The figure also shows the orientation of the Cartesian coordinate system consistently used throughout the text.A speech processing system converts lexical presentation to a string of phonemes and a computer animation system, given visual data, generates immediate frames to provide a smooth transition from one visual image to the next to provide smooth animation. The timing data is used to correlate the phonemic string with the visual images to produce accurately timed sequences of visually correlated events. Utilizing a real time random access audio/visual synthesizer **(RAVE)** together with associated special purpose audio/visual modeling language **(RAVEL)** synthesized actors**(synactor)** representing real or imaginary people, animated characters can be simulated and programmed to perform actions.

MAJOR SPEECH RELATED FACIAL MOVEMENTS

For controlling the dynamic behavior of a 3D synthetic face, the speech related movements are broken down into three components: a rigid body component associated with the movement of the mandible, bounded by the behavior of the **temporomandibular joint,** and two non-rigid components associated with movements of the upper and lower lips. In the following paragraphs, we derive behavioral models for the temporomandibular joint and the lower lip based on the information provided by the trajectories of the fiduciary points. The upper lip behavior is directly given by the trajectory of P1.



Considering the conventions in above fig and the measured coordinates y4(t) and z4(t) of the fiduciary point P4 located on the chin, the TMJ angle of rotation can be expressed by

$$\theta_4(t) = \arctan\left(\frac{z_4(t) - c_z(t_0)}{y_4(t) - c_y(t_0)}\right),$$

where cy(t0) and cz(t0) are the components on the midsagittal plane of $\overline{C}(t_0)$, with the TMJ center at rest position. The presumed location of the TMJ center of our speaker was also measured from the video. The translations ty(t) and tz(t) of the TMJ center within the midsagittal plane can be expressed by

$$\begin{cases} t_y(t) = y_4(t) - c_y(t_0) - r_4 \cos(\theta_4(t)), \\ t_z(t) = z_4(t) - c_z(t_0) - r_4 \sin(\theta_4(t)). \end{cases}$$

The radius r4, the module of vector $\overline{R}_4(t)$, is a constant and is defined by the size of the mandible. It is possible to estimate r4 in the position of rest by

$$r_4 = \sqrt{[y_4(t_0) - c_y(t_0)]^2 + [z_4(t_0) - c_z(t_0)]^2}.$$

LOWER LIP BEHAVIOUR

The lower lip behavior is related to the voluntary movement of the lower lip tissue (other than that caused by the mandible) necessary to produce specific speech gestures, such as lip protrusion. Lower lip behavior describes this kind of voluntary movement involved in the trajectory of the fiduciary point P3. Non-rigid body deformation can be derived from this behavior to control the animation of the lower region of the synthetic face around the mouth.Assuming that $\overline{P}_3(t)$ is the trajectory of the fiduciary point P3 and that $\overline{R}_3(t)$ is the movement of the lower lip due to that of the TMJ, the lower lip behavior can be expressed by

$$\overline{P}_3(t) - \overline{R}_3(t) - \overline{C}(t_0)$$

The components of the vector $\overline{R}_3(t)$ in the midsagittal plane are (directions Y and Z, respectively)

$$\begin{cases} r_{3_y}(t) = r_3\ \cos(\theta_3(t)) + t_y(t), \\ r_{3_z}(t) = r_3\ \sin(\theta_3(t)) + t_z(t). \end{cases}$$

The distance r3 between P3 and the TMJ center can be calculated from the coordinates of P3 and the location of the center of the TMJ at rest by

$$r_3 = \sqrt{[y_3(t_0) - c_y(t_0)]^2 + [z_3(t_0) - c_z(t_0)]^2}.$$

As the lower lip experiences the same variation of angle as P4 due to the rotation of the TMJ, θ3(t) can be expressed by

$$\theta_3(t) = \theta_4(t) - \theta_4(t_0) + \arctan\left(\frac{z_3(t_0) - c_z(t_0)}{y_3(t_0) - c_y(t_0)}\right).$$

These figures show that the analysis of the lower lip behavior correctly indicates the occurrence of lower lip protrusion. This protrusion occurs when the solid curve in the figures is to the left of the dashed one. When it is to the right, the lower lip is retracted, as in mouth spreading.

UPPER LIP BEHAVIOUR

As with lower lip behavior, upper lip behavior is associated with the voluntary movement of lip tissue. However, in contrast to what happens with the lower lip, the behavior of the upper lip associated with the fiduciary point P1 is not entangled with TMJ movement. Thus, upper lip behavior can be determined directly from the trajectory of the fiduciary point P1. Non-rigid body deformation is derived from this behavior to control the animation of region above the mouth of the synthetic face.

VIRTUAL FACE MANIPULATION

To reproduce the movements of the mandible in the synthetic face, the rigid body transformations of rotation and translation described by TMJ behavior are applied to the polygonal vertices of the geometric model. These transformed vertices are those within the region of the face alongside the mandibular bone, more precisely, in the region below and including the lower lip and the lateral side of the face below an imaginary plane defined by TMJ rotation axis and the corners of the mouth. The lower bound of the mandibular region is defined by the neck. The white dots shown in the below fig. are the vertices of the synthetic face submitted to these transformations. The axis of rotation defined by the center of the TMJ is presented in the figure as a white line in front of the ear.The behavior of the upper and lower lips is mapped onto the synthetic face on the basis of three main considerations. First, the points on the geometric model corresponding to the fiduciary points must closely mimic the behaviors described by the measurements. Second, during speech production, the skin tissue around the mouth, including the lips, suffers deformations primarily attributed to the sphincteral behavior of the **Orbicularis Oris** muscle. Third, other muscles which also influence the movement of the skin around the mouth are distributed asymmetrically with respect to the horizontal plane (with different groups of muscles inserted into the skin of the upper and lower regions of the mouth). Based on these considerations, a geometric model was derived to express the visible characteristics of the skin around the mouth during speech production.

VISEME REPRESENTATION

The visual representation of the phonemes, or visemes, is a crucial aspect of speech synchronized realistic facial animation. The recognizable characteristics of a viseme, however, can significantly change due to the effects of coarticulation. Traditionally, these effects have been modelled by fusion functions that blend phonetic context independent visemes. **Context-independent visemes**, however, are “pure” visemes that represent an idealization of expected speech postures. This idealization assumes that the speech segment, and therefore its viseme, is produced in an isolated form, without suffering interference from the articulatory movements of a nearby segment. The characterization of a “pure” viseme and of its blending involves an intricate process and requires the gathering and analysis of extensive articulatory data.

VIDEO TO SPEECH

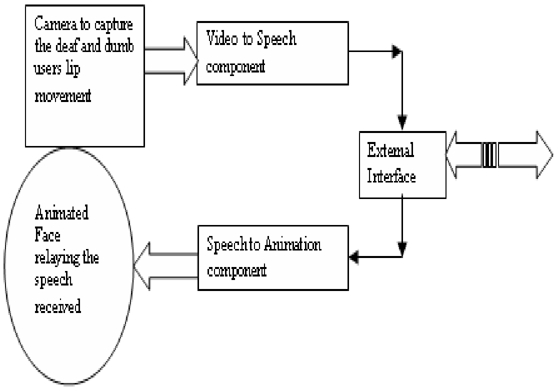
This application shall basically do the reverse of Speech to Animation application. It shall convert live video streams to words, which shall be converted to phonemes and thereby to speech streams that can be relayed over the network. The images of the deaf and dumb user are converted to visemes which are then converted as phonemes and then relayed to the blind user as voice signals.

A TYPICAL HANDSET- A BLIND USER CALLS A DEAF, DUMB USER.

A typical cell phone for deferent echelon of users will have a camera to capture the facial expressions esp. the lip movements of the blind user which are the converted to visemes using viseme representation techniques discussed here. Later geometric models are derived for these visemes and animated face is created using **synchronized facial recognition.** The audio signals of the blind speaker are fused with the lip movements to generate visemes. Such a fusion is called **multi-modal fusion** and the system is said to be a **multimodal speech recognition system.** But for speech recognition, it is usually sufficient to extract the principal components of the lip information and to match the mouth openings–closings with the phonemes of speech.The deaf, dumb user at the receiving end receives the animated images of the transmitted information. The camera in his cell phone captures his lip movements and converts them to voice signals by reversing the process , thus enabling the blind user to hear the voice messages.

A SAMPLE PHONE CAL: BLIND USER(X) TO DEAF, DUMB USER(Y)

- X issues voice command to call Y
- Y receives a vibrating alert.
- Upon accepting the call, the animated face opens up in Y’s cell phone
- Y’s response is captured by the camera and is streamed to video to speech converter, which is then transmitted to X.
- The camera in Y’s phone is ready as well.
- X’s response (speech) lands in Y and is converted to animation stream which is played by the face in Y. For speech recognition, it is usually sufficient to extract the principal components of the lip information and to match the mouth openings–closings with the phonemes of speech.


Camera to capture the deaf and dumb users lip movement
Video to Speech component
External Interface
Speech to Animation component
Animated Face relaying the speech received

ADVANTAGES AND CHALLENGES

This recommendation is quiet effective to because of many factors. The main, being the availability of higher end technology that already supports these (STA) applications. Most of the applications are language independent or multilingual in other words. Some applications provide greater authentic facial movement replication than a normal eye could translate, thus rendering the application more effective. Above all, Lip reading is the most effective way of communication used by the deaf and dumb. The blind person could use a lesser complicated gadget or even closer to the normal gadget.

On the flip side,

- Not all deaf and dumb users can lip read. A deaf and dumb person by birth has a lesser possibility to posses the skill of lip reading. Ideally, the user should be trained in lip reading.
- The video to speech converter application using phonemes is still not common.
- The characterization of a “pure” viseme and of its blending involves an intricate process. Moreover the implementation of this model using poorly defined parameters can easily produce deceptive results.

CONCLUSIONS

This recommendation is highly effective in many ways. The purpose of science and technology esp. wireless communication is that it should reach the masses. And as communication engineers it is our forthright duty to extend the usage of wireless technology to the impaired also. The technologies mentioned in this paper exist already and it would very fruitful if the suggested plan is implemented. It would help the deferent sector of the society to keep abreast of the developments in science and technology.

"Any sufficiently advanced technology is indistinguishable from magic. For man holds in his mortal hands the power to abolish all forms of human poverty, and all forms of human life."

-John Fitzgerald Kennedy

REFERENCES

[1] C. Pelachaud, N.I. Badler and M. Steedman, Generating facial expressions for speech, Cognitive Science

[2] P.L. Jackson, The theoretical minimal unit for visual speech perception: Visemes and coarticulation

[3] A.-P. Benguerel and M.K. Pichora-Fuller, Coarticulation effects in lipreading, Journal of Speech and Hearing Research

[4] N.P. Erber, Auditory, visual, and auditory-visual recognition of consonants by children with normal and impaired hearing, Journal of Speech and Hearing Disorders

[5] I. Matthews, T. Cootes, J. Bangham, S. Cox and R. Harvey, Extraction of visual features for lipreading

[6] G. Potamianos, C. Neti, G. Gravier, A. Garg, A. Senior, Recent advances in the automatic recognition of audio-visual speech.