Intelligent Design for Sound Detection of Goblet Clinking

Min-Xiang Huang[1], Ho-En Liao[2], Ming-Hwa Sheu[1], Jia-Xiang Zheng[1], Yu-Cheng Ou[1], Tzu-Hao Kuo[1]

[1]*Department of Electronic Engineering, National Yunlin University of Science & Technology, Yunlin, Taiwan*
[2]*Department of Communications Engineering*, *Feng Chia University, Taichung, Taiwan*

***Abstract*-- With the development of Internet of Things (IoT) lately, various applications of communications among objects rapidly prevail over our life. This work aims to explore an application of IoT that is capable transforming colors of goblets to add extra fun or ambiance in social occasions. To achieve this work, we write a mobile phone APP to detect toast clink through a sound sensor and to drive LED modules embedded at the bottoms of goblets through Bluetooth. This work can be further developed to have more vivid color transforming of goblets by detecting, such as, loudness of clink, amount of liquid left, speed of drinking, and so on via a variety of sensor modules.**

I. INTRODUCTION

The use of drinks as a social modality is prevailed over different cultures or countries around the world, and it is common to have goblets as social tools. Moreover, to greet or bless with enthusiasm, or to have closer relationship among people, we gently knock or clink goblets to make sound of clink. In order to increase the fun or ambiance of drinking, there are related works such as [1] and [2]. In [1], cups are designed to be capable of producing sound to entertain children while drinking. On the other hand, [2] dedicates to designing the structure of a gleamy goblet with a plurality of tiny color bulbs with pre-designated embedded.

In this work, we configure a mobile phone based IoT system that can transform the color of each goblet individually by LED illuminance from the bottom or stem of the goblet. The system use the microphone of a mobile phone to correct acoustic sound in environment, and detect clink sound generated by goblets. Once a clinking sound is recognized, a wireless device will be triggered to transmit identification codes and color data to each goblet site and, thereafter, each LED module embedded in a goblet will be commanded to illuminate color as designated. Note that the color presented can be different from each other and time-alternated. The wireless device we used is a Bluetooth transceiver module since it is less power consumption and is able to transmit data to several clients. The most critical subject of this work is to develop an algorithm that can correctly pick up goblet clinking sound in noisy environment so that LED's will illuminate at the right moment. To detect and recognize the clinking sound, there are three phases of processing. The first phase compute the use zero-crossing rate (ZCR) of a sound frame to identify if it is merely noise. If a sound frame was identified as "not noise," then we could go forward to the second phase to compute Mel-scale Frequency Cepstral Coefficients (MFCC) so that a "sound signal" can be extracted. MFCC is composed of a sequence of processes including pre-emphasis, high-pass filtering, Hamming windowing, FFT and DCT. Thereafter, the third phase compares MFCC of sample clinking sound with that of sound extracted by normalized cross-correlation to determine if a goblet clinking sound exists.

In the following, we illustrate the use of MFCC to extract sound signal and clinking sound recognition in Section II. In Section III, we illustrate the architecture of the work. Conclusions are made in Section IV.

II. PROPOSED METHODOLOGY

A. Mel-scale Frequency Cepstral Coefficients

Mel-scale Frequency Cepstral Coefficients (MFCC) takes into account the reception of human ears, and is therefore a more mature method to extract speech feature. Hence MFCC method is widely applied in the field of speech recognition as a pre-processing phase to extract speech signal, e.g., [3] [4] [5]. Fig.1 shows the block diagram of MFCC, and each block is explained in follows.

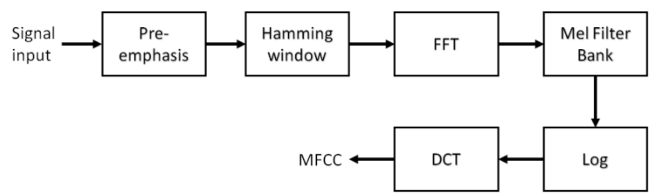


Fig. 1 MFCC block diagram

i. Pre-emphasis

The voice signal s(n) is fed to a high-pass filter with transfer function, H(z),

$$H(z)=1-a*z-1 \quad (1)$$

, where a is constant between 0.9 and 1.0. If we express (1) in time domain, then the pre-emphasized signal $s_1(n)$ is

$$s_1(n) = s(n) - a * s(n-1).$$

The purposes of this pre-emphasis are to eliminate the effects of vocal cords and lips when generating sound and, on the other hand, to compensate the suppressed high-frequency components.

ii. Hamming window

The purpose of the Hamming windowing is to make a frame of sound signal continuous, i.e., to eliminate discontinuities, so that the distortion generated at the FFT computation can be lowered. The Hamming window formula is shown in (2), where n is the current sound point, N is the number of points within a sound frame, and α is set to 0.46 in our work.

$$W(n) = (1 - \alpha) - \alpha * \cos\left(\frac{2\pi n}{N-1}\right) \quad (2)$$

iii. Fast Fourier Transform (FFT)

Because the characteristics of sound waveform in time-domain is not clear to analyze, it is usually have the waveform transform into frequency domain so that we can have the magnitude distribution of frequency components. This make us be able to distinguish different voice much easier.

iv. Mapping of the spectrum to Mel scale

To flatten the harmonics of a speech signal and extract the corresponding envelope after FFT, a set of triangular band-pass filters that are equally spaced along the Mel-frequency are used as shown in (3). This makes a speech recognition system have similar results when inputs are of the same timbre but with different pitch.

$$\mathrm{mel(f)} = 2595 * log_{10}(1 + \frac{f}{700}) \quad (3)$$

v. Logarithmic conversion

To accommodate the characteristics of the sound response of human ears, each sound frame of the Mel filter bank output is transformed to logarithmic magnitude. This can reduce the numerical dynamic range and calculate the loudness or intensity of a sound frame.

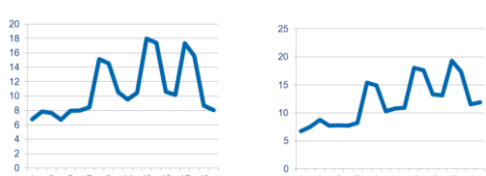
vi. Discrete Cosine Transform

Since the IFFT of the coefficients generated by the log-Mel filter may be complex and asymmetric, DCT is applied to determine the cepstral coefficients.

B. Normalized Cross-Correlation Identification

We use the waveform generated by MFCC as a sample to compare it with the actual incoming sound to determine if there exists a clinking sound. The method of comparison uses the normalized cross-correlation coefficient, as shown by (4), to determine whether the incoming sample is similar to the standard one. As the magnitude of the cross-correlation coefficient is closer to unit, the degree of similarity between the two waveforms is higher. And, a threshold is pre-assigned to determine if the incoming waveform is the same class as the sample one. Fig.2(a) and (b) show a sample of goblet clinking sound and an actually incoming clink feature, and the normalized cross-correlation coefficient calculated is 0.9858. This shows high similarity between the two waveforms.

$$Ncc(d) = \frac{C(i,j)}{[\sum_{x=0}^{L-1}\sum_{y=0}^{20} w(x,y)^2]^{1/2}[\sum_{x=0}^{L-1}\sum_{y=0}^{K-1} f(x+i,y+j)^2]^{1/2}} \leq 1 \quad (4)$$



(a) (b)

Fig.2 (a) Goblet clinking sound feature of the sample (b) Goblet actual clinking feature

III. Architecture Design

Our platform use mobile phone with Android system as master and use wireless Bluetooth transceiver to communicate among goblets. Whenever a clinking sound is identified by the master, color information is transmitted to have LED's illuminate the color as designated. The system flow chart is shown in Fig. 3. Fig. 4(a) shows the possible

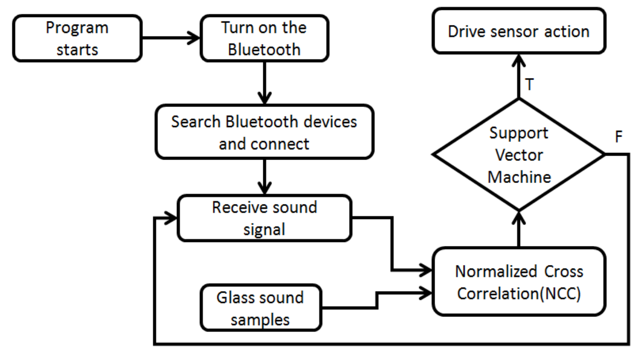
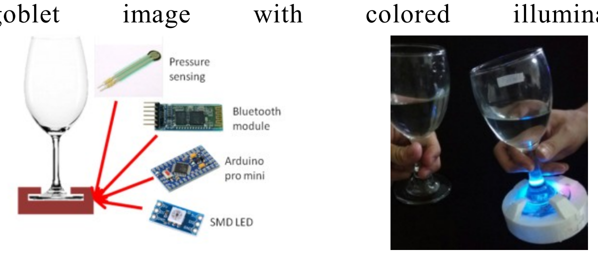


Fig. 3 System flow chart

modules used in slave (goblet) site, and Fig. 4(b) shows a goblet image with colored illumination.



(a) (b)

Fig.4 (a) Modules at goblet site; (b) colored goblet image

IV CONCLUSION

In this paper, we use MFCC and normalized cross-correlation coefficient to identify clinking sound. We need to have a training sample so that comparison Future works may lead more advanced AI technology to have more accurate sound identification and design an embedded system to integrate modules required to fit the size for practical situation.

REFERENCES

[1] Yuen Wing Lo, Chi Hing Kwong, " SOUNDING CUP ," United States Patent Application Publication, Pub. No.: US 2001/0041945 A1, Pub. Date: Nov. 15, 2001.

[2] Chiu-Tung Chen, " **Gleamy goblet structure**," United States Patent, Pub. No.: US 2001/0041945 A1, Pub. Date: Nov. 15, 2001.

[3] Achmad Lukman, Agus Harjoko, Chuan-Kay Yang, "Classification MFCC feature from Culex and Aedes aegypti Mosquitoes Noise using Support Vector Machine ," International Conference on Soft Computing, Intelligent System and Information Technology, Denpasar, Bali, Indonesia, Sept. 2017, pp.17-20.

[4] A. Maazouzi, N. Aqili, A. Aamoud, M. Raji, A. Hammouch, "MFCC and Similarity Measurements for Speaker Identification Systems ," International Conference on Electrical and Information Technologies, Rabat, Morocco, Nov. 2017, pp.1-4.

[5] M.S. Nagawade, V.R. Ratnaparkhe, "Musical Instrument Identification using MFCC ," 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology , May 2017, pp. 2198 - 2202.