Human Voice Based Consumer Electronic Device Operating System

M. Bhuvaneswari[1], K. Praveen Kumar[2]

M.E. (Embedded System Technologies), Mahath Amma Institute of Engineering and Technology, Pudukkottai, Tamil Nadu, India.

Head of the Department, Dept. of ECE, Mahath Amma Institute of Engineering and Technology, Pudukkottai, Tamil Nadu, India.

ABSTRACT: A system is proposed that can remotely operate consumer electronic devices by voice. It uses the mobile phone as a controller. And it uses the CELP(code excited linear prediction) parameters that are used for speech coding in mobile phones. A speaker verification function protects private information and separates the user's voice from that of people nearby who are also speaking. A CELP-based speaker verification method is used to match the audio stream by comparing the trajectories of continuous phonemes. Experimental evaluation of the speaker verification method demonstrated the effectiveness of the proposed verification method.

KEYWORDS: consumer electronic device, mobile phone, CELP, LSP, speaker verification.

I. INTRODUCTION

The operation of devices by voice is receiving increasing attention because operation by voice is easier than by keyboard input or screen touching. The voice signals are transformed into commands by speech recognition. Another area receiving increasing attention is the "Internet of things" (IoT) [1], [2], in which consumer electronic devices such as TVs, air conditioners, refrigerators, hard-disk recorders, and home appliances are connected to the Internet. Moreover, consumer electronic devices can be controlled remotely by using a mobile phone or other terminal. For example, a home's air conditioner can be turned on by using one's mobile phone before reaching home on a hot day. Such operations are more convenient when done using voice, as illustrated in Fig. 1. However, some consumer electronic devices contain private information such as usage history, user profile, and archived contents. If someone who finds or steals a mobile phone uses it to control such devices, that person may be able to access the owner's private information.

Therefore, the person using a mobile phone to control consumer electronic devices should be verified. Another problem with using a mobile phone to control consumer electronic devices by voice remotely is that more than one person may speak at the same time. This can result in unintentional device operation. Therefore, the speaker's identity should be verified, and the user's instructions should be separated from the voice signals of other persons. That is, there is a need to know who is speaking at each moment. Therefore, it is necessary to verify the input voice. Given this need for user verification, a verification function should be implemented in consumer electronic devices.
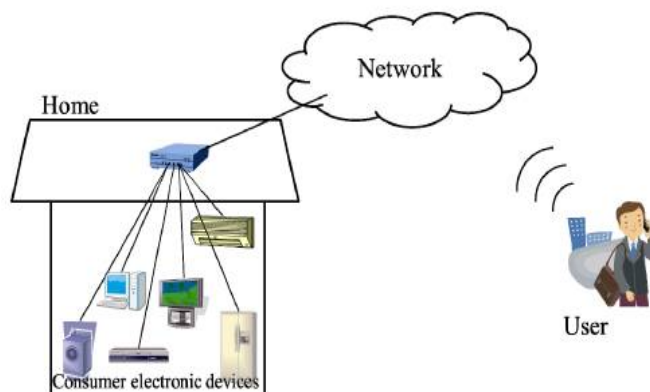
Fig. 1 Control of consumer electronic devices from outside the home

A variety of people use consumer electronic devices, and it is important that users be able to use them easily. Voice operation is particularly useful. It is also important that remote operation be versatile. Operation by voice can be supported in various places if the mobile phone holder's voice is separated from the speaking of people nearby who are also speaking. Furthermore, a speaker verification function to be implemented in consumer electronic devices should not be expensive. Moreover, it should be possible to operate such devices by voice without requiring a special sensor or equipment.

That is, encoders, decoders, and other sensors should not be required for implementing a speaker verification function. Speaker verification can be done using the parameters that are encoded in mobile phones. An encoder and microphone for capturing voice are typically installed in mobile phones. A system is proposed here for operating consumer electronic devices by voice using CELP (code excited linear prediction) parameters, which are commonly used in speech coding for mobile phones. CELP is an AbS (Analysis by Synthesis) coding scheme with an excited sound source which is prepared in a codebook. Also proposed is a method for verifying the speaker's identity by using CELP parameters. Experimental evaluation of the speaker verification method demonstrated the effectiveness of the proposed method.

PROPOSED SYSTEM FOR OPERATING CONSUMER ELECTRONIC DEVICES BY VOICE REMOTELY

The system proposed for operating consumer electronic devices by voice remotely was designed to meet several requirements.

- It should support easy remote operation of consumer electronic devices.
- It should be able to operate in various places without execution of unintentional instructions.
- It should operate without the use of other equipment.
- It should verify the user's identity to prevent use by another person.

The proposed system supports voice operation by using the CELP parameters that are encoded in mobile phones. As illustrated in Fig. 2, CELP parameters are used in the speaker verification and speech recognition functions, which are assumed to be performed on consumer electronic devices to prevent misuse by a malicious user.

II.   RELATED WORK

RELATIONSHIP BETWEEN THE INTERNET OF THINGS AND CONSUMER ELECTRONICS

Few things can be more exciting than introducing a new column to a large group of respected professionals, especially when it is in the flagship IEEE Consumer Electronics Magazine of a renowned Society. Here, I am thrilled to introduce to you the new column—"Internet of Things." "Internet of things," usually abbreviated as IoT, is a popular term that has been widely spread and adopted in the past years since it was first used [1]. Wikipedia gives its definition as "uniquely identifiable objects and their virtual representations in an Internet-like structure" [2]. The European IoT Hub [3] proposes a more aggressive definition as "an integrated part of the future Internet" and "a dynamic global

network infrastructure… where physical and virtual things… are seamlessly integrated into the information network." Actually, in the rapidly growing and expanding practices spread across Asia, Europe, and other continents, IoT refers to any systems of interconnected people, physical objects, and IT platforms, as well as any technology to better build, operate, and manage the physical world via pervasive data collection, smart networking, predictive analytics, and deep optimization. As you can see, the concept of IoT is very relevant to, if not known as, machine-tomachine and cyberphysical systems, which are popular terms.

ANALYSIS OF REMOTE CONTROL TECHNIQUES EMPLOYED IN HOME AUTOMATION AND SECURITY SYSTEMS

Remote control of home appliances and remote alerting on fire detection and intruder detection in home are essential aspects of home automation and home security systems. Most commonly used communication tools such as cellular phones, landline telephone and email are considered for remote control of appliances by proposing efficient structures. All these control schemes are analyzed for their performances suiting modern home automation and security demands. Furthermore, with the installation of sensitive fire detector and intruder detector in home, automatic triggering of a unit sending recorded message converted into email is presented. This aspect strengthens further the security concerns of the home.

HANDS-FREE VOICE COMMUNICATION WITH TV

This paper presents a system for full-duplex hands free voice communication integrated with TV technology. The system provides comfortable conversation by utilization of microphone array and advanced voice processing algorithms, even with simultaneous TV usage. Signal processing includes superdirective beamformer steered by direction-finding module, postprocessing module, acoustic echo canceller, stationary noise reduction module and automatic gain control. All processing is realized in real-time on DSP based platform. As communication channel GSM or VoIP can be used.

INTRODUCING BLUETOOTH SMART: PART II: APPLICATIONS AND UPDATES

In 2010, the bluetooth special interest group published their Core Specification 4.0, including Bluetooth Low-Energy (BLE) technology, or Bluetooth Smart. This short-range radio breaks technical ground: it can run for years on batteries or scavenged power. It is a leading candidate to connect the Internet of Things.? This was updated to Core Specification v4.1 in late 2013 [1]. For more details, including reasons for decisions, see Bluetooth Low Energy, The Developers Handbook.

SPEECH VERSUS KEYING IN COMMAND AND CONTROL: EFFECT OF CONCURRENT TASKING

As a result of Poock's influential work in the early 1980s, command and control is generally believed to be one specific application where speech input holds great advantages over keyed data entry. However, a recent paper (Damper & Wood, 1995 "Speech versus keying in command and control applications ",International Journal of Human-Computer Studies,42,289-305) has questioned this interpretation of Poock's data because the experimental conditions seemed to bias the results against keyed entry. While Damper and Wood modelled their experiments on Poock's, however, there were important differences which mean that their conclusions are uncertain. The objective of the work reported here was to determine if the major difference-the omission of concurrent, secondary tasking from their study-could explain Damper and Wood's observed superiority of keying over speech. Simulated command and control experiments are described in which speech input, abbreviated command keying and full command keying are compared under dual-task conditions. We find that speech input is no faster (a non-significant 1.23% difference) and enormously more error-prone (1038%, highly significant) than abbreviated keying for the primary data entry task, but allows somewhat more (11.32%, not significant) of a secondary information-transcription task to be completed. Full keying has no advantages whatsoever: we believe that this confirms the methodological flaw in Poock's work. If recognizer errors (as opposed to speaker errors) are discounted, however, speech shows a clear superiority over keying. This indicates that speech input has potential for the future-especially for high workload situations involving concurrent tasks-if the technology can be developed to the point where most errors are attributable to the speaker rather than to the recognizer.

KERNEL MUTUAL SUBSPACE METHOD AND ITS APPLICATION FOR OBJECT RECOGNITION

A multiple observation-based scheme (MObS) is described for robust facial recognition, and a novel object

recognition method called kernel mutual subspace method (KMS) is proposed. The mutual sub-space method (MSM) proposed by (Maeda, et al., 1999) is a powerful method for recognizing facial images. However, its recognition accuracy is degraded when the data distribution has a nonlinear structure. To overcome this shortcoming we apply kernel principal component analysis (kPCP) to MSM. This paper describes theoretical aspects of the proposed method and presents the results of facial image recognition experiments

HANDS-FREE VOICE COMMUNICATION WITH TV - I. I. PAPP, Z. M. SARIC

This system presents a system for full-duplex hands free voice communication integrated with TV technology. The system provides comfort conversation by utilization of microphone array and advanced voice processing algorithms, even with simultaneous TV usage. As communication channel GSM or VoIP can be used. Hands-free voice communication platform integrated with TV.

ACOUSTIC INTERFERENCE CANCELLATION FOR A VOICE-DRIVEN INTERFACE IN SMART TVS - J. S. PARK, G. J. JANG, J. H. KIM

A novel method is proposed to improve the voice recognition performance by suppressing acoustic interferences that add nonlinear distortion to a target recording signal when received by the recognition device. The proposed method is expected to provide the best performance in smart TV environments, where a remote control collects command speech by the internal microphone and performs automatic voice recognition, and the secondary microphone equipped in a TV set provides the reference signal for the background noise source. Due to the transmission channel, the original interference is corrupted nonlinearly, and the conventional speech enhancement techniques such as beamforming and blind signal separation are not applicable. The proposed method first equalizes the interference in the two microphones by maximizing the instantaneous correlation between the nonlinearly related target recording and reference signal, and suppresses the equalized interference. To obtain an optimal estimation of the equalization filter, a method for detecting instantaneous activity of interference is also proposed. The validity of the proposed method is proved by the improvement in automatic voice recognition performance in a simulated TV room where loud TV sounds or babbling speech interfere in a user's commanding speech.

SPEECH VERSUS KEYING IN COMMAND AND CONTROL: EFFECT OF CONCURRENT TASKING - R. I. DAMPER, M. A. TRANCHANT AND S. M. LEWIS

As a result of Poock's influential work in the early 1980s, command and control is generally believed to be one specific application where speech input holds great advantages over keyed data entry. However, a recent paper (Damper & Wood, 1995 "Speech versus keying in command and control applications ",International Journal of Human-Computer Studies,42,289-305) has questioned this interpretation of Poock's data because the experimental conditions seemed to bias the results against keyed entry. While Damper and Wood modelled their experiments on Poock's, however, there were important differences which mean that their conclusions are uncertain. The objective of the work reported here was to determine if the major difference-the omission of concurrent, secondary tasking from their study-could explain Damper and Wood's observed superiority of keying over speech. Simulated command and control experiments are described in which speech input, abbreviated command keying and full command keying are compared under dual-task conditions. We find that speech input is no faster (a nonsignificant 1.23% difference) and enormously more error-prone (1038%, highly significant) than abbreviated keying for the primary data entry task, but allows somewhat more (11.32%, not significant) of a secondary information-transcription task to be completed. Full keying has no advantages whatsoever: we believe that this confirms the methodological flaw in Poock's work. If recognizer errors (as opposed to speaker errors) are discounted, however, speech shows a clear superiority over keying. This indicates that speech input has potential for the future-especially for high workload situations involving concurrent tasks-if the technology can be developed to the point where most errors are attributable to the speaker rather than to the recognizer.

SPEECH VERSUS TOUCH: A COMPARATIVE STUDY OF THE USE OF SPEECH AND DTMF KEYPAD FOR NAVIGATION - K. M. LEE AND J. LAI

Speech vs. Dual Tone Multiple Frequency [DTMF] in a phone-based message retrieval system. Unlike previous studies that compared these two modalities, the speech system used in this study was a fully functioning natural language system, and participants in this study were working professionals, rather than college students. Results indicate that (a) DTMF was more effective and efficient for linear tasks, whereas speech was better for nonlinear tasks;

(b) speech was preferred to DTMF by a majority of users; (c)speechwasjudgedasbeingmoresatisfying,moreentertaining,andeasiertousethan DTMF; and (d) user preference for a particular modality was better predicted by user performance in nonlinear tasks rather than linear ones. Possible reasons for users ' continuing preference for the speech modality even after experiencing fairly high recognition errors are discussed. Finally, the importance of examining speech user interfaces from other perspectives, in addition to efficiency maximization, is emphasized. The results of this study have theoretical, as well as practical, implications for the design of speech user interfaces and interactive voice response applications.

CELP-BASED SPEAKER VERIFICATION: AN EVALUATION UNDER NOISY CONDITIONS - Y. YAMAZAKI, Y. FUJITA, AND N. KOMATSU

We propose a text-independent speaker verification method based on a speech coding scheme. The proposed method utilizes CELP parameters which are used in speech coding schemes for mobile communication systems, and verifies a speaker only with the encoded speech information. The reliability of the proposed method under noisy conditions is mainly discussed with some simulation results.

SPEAKER RECOGNITION USING G.729 SPEECH CODEC PARAMETERS - T.F. QUATIERI, R.B. DUNN, D.A. REYNOLD, J.P. CAMPBELL

Experiments in Gaussian-mixture-model speaker recognition from mel-cepstra, derived from mel-filter bank energies (MFBs) of the G.729 codec all-pole spectral envelope, showed significant performance loss relative to the standard mel-cepstral coefficients of G.729 synthesized (coded) speech (Quatieri et al. 1999). In this paper, we investigate two approaches to recover speaker recognition performance from G.729 parameters. The first is a parametric approach that makes explicit use of G.729 parameters, rather than deriving cepstra from MFBs of an all-pole spectrum. Specifically, the G.729 LSFs are converted to "direct" cepstral coefficients for which there exists a one-to-one correspondence with the LSFs. The G.729 residual is also considered; in particular, appending G.729 pitch as a single parameter to the direct cepstral coefficients gives further performance gain. The second nonparametric approach uses the original MFB paradigm, but adds harmonic striations to the G.729 all-pole spectral envelope. Although obtaining considerable performance gains with these methods, we have yet to match the performance of G.729 synthesized speech, motivating the need for representing additional fine structure of the G.729 residual.

III. CONCLUSION

The system to remotely operate consumer electronic devices by voice using the CELP parameters that are used for speech coding in mobile phones was proposed. A speaker verification function protects private information and separates the user's voice from that of people nearby who are also speaking. A CELP-based speaker verification method is used to match the characteristics of a set of frames by comparing the trajectories of continuous phonemes. Experimental evaluation of the verification method demonstrated its effectiveness. A method based on CELP parameters for separating the user's voice from that of people nearby who are also speaking will be studied in future work.

REFERENCES

[1] Y. Yuan, "Relationship Between the Internet of Things and Consumer Electronics," IEEE Consumer Electronics Magazine, p.23, Apr. 2012.

[2] J. Decuir, "Introducing Bluetooth Smart: Part II: Applications and updates," IEEE Consumer Electronics Magazine, pp.25-29, Apr. 2014.

[3] J. Han, J.K. Yun, J.H. Jang and K.R. Park, "User-Friendly Home Automation Based on 3D Virtualb World," IEEE Trans. on Consumer Electron., pp.1843-1847, Aug. 2010.

[4] K. Balasubramanian and A. Cellatoglu, "Analysis of remote control techniques employed in home automation and security systems," IEEE Trans. on Consumer Electron., Vol.55, Issue 3, pp.1401-1407, Aug.2009.

[5] T. Kim, H. Lee and Y. Chung, "Advanced Universal Remote Controller for Home Automation and Security," IEEE Trans. on Consumer Electron., pp.2537-2542, Vol. 56, Issue 4, Nov. 2010.

[6] Y.R. Chuang, W.J. Yang S.J. Lin and T.L. Chiu, "Study and implementation of the smallest closed-area (SCA) mechanism for selforganization network architectures in smart home control systems," IEEE International Symposium on Consumer Electronics, pp.79-80, Jun. 2013.

[7] I. I. Papp, Z. M. Saric and N.D. Teslic, "Hands-free Voice Communication with TV," IEEE Trans. on Consumer Electron., pp.606-614, Vol. 57, Issue 2, May 2011.

[8] J. S. Park, G. J. Jang, J. H. Kim and S. H. Kim, "Acoustic Interference Cancellation for a Voice-driven Interface in Smart TVs," IEEE Trans. on Consumer Electron., pp.244-249, Vol. 59, Issue 1, Feb. 2013.

[9] R. I. Damper, M. A. Tranchant and S. M. Lewis, "Speech versus keying in command and control: effect of concurrent tasking," International Journal of Human-Computer Studies, pp.337-348, Vol. 45, issue 3, Sep. 1996.

[10] K. M. Lee and J. Lai, "Speech Versus Touch: A Comparative Study of the Use of Speech and DTMF Keypad for Navigation," International Journal of Human-Computer Interaction, pp.343-360, Vol. 19, issue 3, Jan. 2005.

[11] Y. Yamazaki, Y. Fujita, and N. Komatsu, "CELP-based Speaker Verification: An Evaluation under Noisy Conditions," IEEE Internation conference on Control, Automation, Robotics and Vision, pp. 408-412, Dec. 2004.

[12] T.F. Quatieri, R.B. Dunn, D.A. Reynold, J.P. Campbell and E. Singer, "Speaker recognition using G.729 speech codec parameters," IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 1089-1092, Jun. 2000.

[13] H. Sakano, N. Mukawa and T. Nakamura, "Kernel Mutual Subspace Method and Its Application for Object Recognition," Electronics and Communications in Japan, Vol.E88, No.6, pp. 45-53, Jun. 2005.