Glottal Doppler Radar System and Its Applications to Communication and Speaker Recognition

Chia-Chan Chang[1], Chien-San Lin, Sheng-Fuh Chang[3], Chun-Chi Lin,

Zhen-Qiang Jiang and Sung-Nien Yu

Department of Electrical Engineering, Department of Communications Engineering, Center for Telecommunication Research, National Chung Cheng University, Chia-Yi, 621, Taiwan

[1]ccchang@ee.ccu.edu.tw

[3]ieesfc@ccu.edu.tw

***Abstract*—A glottal radar system has been successfully developed to extract human speech information with great immunity to acoustic interference, which enables numerous speech related applications. This work proposed and demonstrated two potential uses. First, for the communication quality enhancement of mobile handsets in a noisy environment, the background acoustic noise can be effectively detected and eliminated based on the glottal radar detection of the unvoiced segments. Second, for the use in speaker recognition, the identification rate of the microphone detection system drops to 82% even though the system is in a 15-dB SNR condition. In contrast, the glottal signal radar still remains 95% identification accuracy.**

I. Introduction

The speech contains both the message being spoken and the information of the speaker, which is the most effective way for human communication. For that reason, scientists and engineers have contributed significantly during past decades in speech processing, such as speech synthesizing, coding and recognition, speaker identification and verification, etc. Most techniques are based on the data of spoken voices. However, under the condition with great amount of background noise, the acoustic signals could be severely degraded due to the imposed interference, which consequently affects the performance of those related techniques, such as communication quality or speech recognition accuracy.

In the recent years, on the other hand, the acoustic speech becomes popular in biometric security system due to its convenience and easy access. Still, the identification accuracy is susceptible to the surrounding noise. Additionally, it is easy to thieve and duplicate the voice signal, which may harm the security.

It has been reported that the periodic motions of vocal cords, can be detected by wireless sensing techniques based on the Doppler radar principle, in real-time, as speech is generated [1]-[4]. The impinging electromagnetic wave is phase modulated due to the periodic displacements of speech organs, more specifically vocal cords. Since the glottal motions provide essential information associated with phonation, it can be treated as one kind of biological signatures. Most importantly, it will provide excellent immunity to background acoustic noise. Those glottal data, combined with corresponding acoustic data, can enable more robust and secure methods in speech related applications.

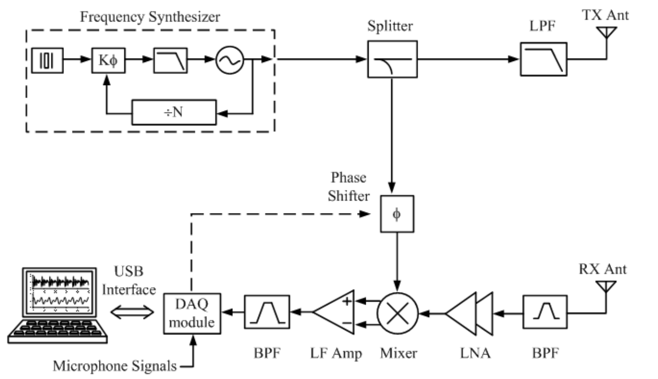

Fig. 1. Block diagram of the proposed glottal radar system (after ref[5]).

A 925-MHz glottal radar system with high sensitivity has been successfully developed in our previous work [5], in which the thorough analysis and system design were addressed in details. Experiments on the detection of vowel and word phonation were also conducted.

The current paper is aimed to investigate the potential applications based on this glottal radar system. The first one is the background acoustic noise reduction, which can potentially be applied in mobile handsets to improve communication quality. In a general scenario, a mobile phone user needs to raise the volume in a noisy scene. The glottal radar can turn off the microphone when it detects the unvoiced period of a speech, thus reducing the background noise in this duration. The second application is for highly-secure speaker recognition. Later the demonstration shows that the glottal radar provides more satisfactory identification accuracy in noisy environment comparing to the one from acoustic signals.

II. System Description

This section briefly outlines the design of proposed glottal radar system. The radar system is realized by a coherent homodyne transceiver, where the circuit block diagram is shown in Fig.1. A phase-locked-loop frequency synthesizer generates a 2-mW, 925-MHz continuous-wave signal. The signal is partially fed back through a phase shifter to the receiver, whereas the rest part is filtered by a low-pass filter for harmonic suppression before being radiated via a transmitter antenna. As the transmitted wave impinges the vibrating glottis, the phase-modulated signal is back-scattered and intercepted by the receiver antenna. The received signals, including the desired phase-modulated signal and other unwanted clutters, are filtered by a band-pass filter to eliminate out-of-band interferences. The filtered signal is then amplified by the low-noise amplifier and coherently down-converted into the baseband. In the baseband domain, a low-frequency amplifier followed by an anti-aliasing filter optimizes the signal-to-noise ratio for driving the followed data acquisition module (DAQ). The DAQ not only executes data computation in digital domain, but also controls the phase shifter to obtain maximal baseband signal intensity. The implemented speech radar system is illustrated in Fig.2. The detailed descriptions on each circuit design have been given in [5].

Since the glottal displacement is extremely small, it is very challenging to achieve remote detection. Currently the antennas should be slightly placed on the larynx area of the human subject for signal detection, as shown in Fig.3. Therefore, this system is very sensitive to the attached position. Any displacement of the antenna could result in variation of the recorded date. In addition to the glottal radar system, an acoustic microphone system is also incorporated to receive the vocal speech, which is used for further data processing or comparison with the electromagnetic glottal signal.

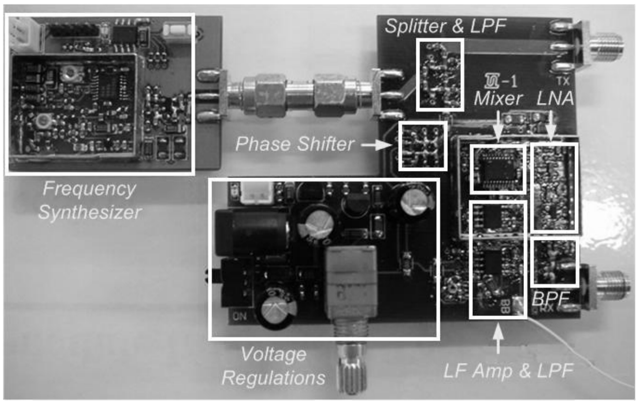

Fig. 2. The implemented glottal radar system (after ref[5]).

Fig. 3. Experimental setup of simultaneous glottal and acoustic signal detection, where the antennas is slightly placed on the larynx area (after ref[5]).

III. APPLICATION FOR ACOUSTIC NOISE REDUCTION

A potential application of glottal radar system is for voice quality enhancement. As shown in Fig. 4, assuming that a mobile handset has build-in glottal radar circuitry, the microphone-captured acoustic signal can be pre-processed by the acoustic noise reduction based on the detected glottal vibration signal.

Figure 5 (a) shows the recorded noiseless acoustic signal, where the test database is "*Jason*". The corrupted acoustic signal is synthesized by the noiseless waveform imposed with -9 dB white noise, shown in Fig. 5(b). Meanwhile, the glottal radar signal, as seen in Fig.5(c), is recorded simultaneously during the phonation. Since the unvoiced period (shadowed areas) can be determined from the glottal signal, the unvoiced segments of the corrupted acoustic signal can be directly discarded and is filtered by a frequency response based on the spectra of glottal radar signal. By applying such a noise reduction process, the corruption noise within the silent periods can be eliminated, as illuminated in Fig. 5(d).

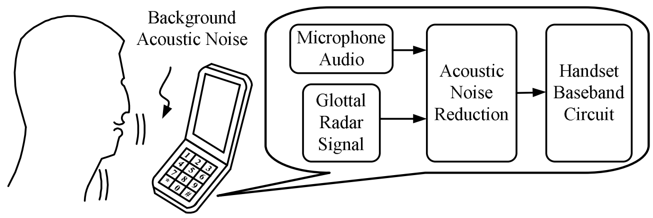

Fig. 4. Illustration of acoustic noise reduction in a mobile handset.

For further analysis, the correlation coefficients for noiseless acoustic signal S1 and the noise-corrupted signal S2 with and without noise reduction are also calculated by (1), as plotted in Fig. 6.

$$\textit{correlation coefficient} = \frac{dot\ (S_1, S_2)}{|S_1| \times |S_2|} \qquad (1)$$

It is obviously that, the correlation coefficient is reduced to 0.71 for the corrupted signal with 6-dB SNR, while it still remains 0.93 when the noise reduction is performed. This successful demonstration reveals the potential enhancement in speech recognition or synthesis by the cooperation between acoustic and glottal radar signals.

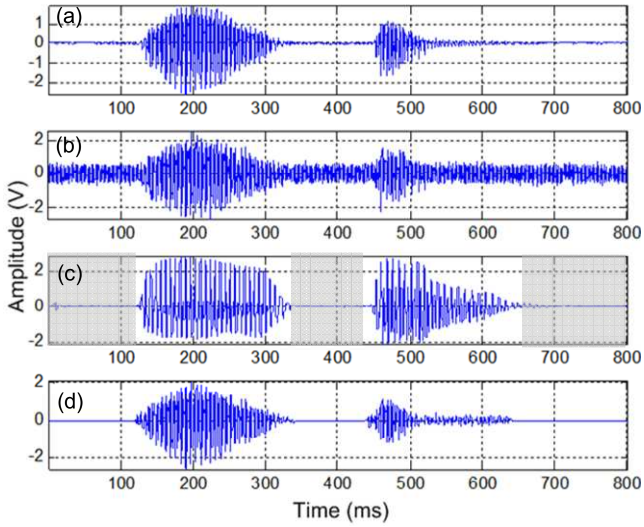

Fig. 5. Experiment results of acoustic noise reduction, (a) noiseless acoustic signal, (b) corrupted acoustic signal with -9 dB while noise, (c) glottal radar signal, (d) reconstructed acoustic signal after noise reduction.

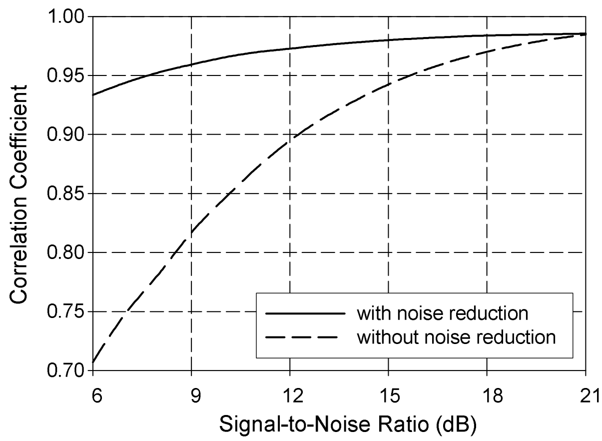

Fig. 6. Comparison of correlation coefficients with and without noise reduction.

IV. APPLICATION OF SPEAKER RECOGNITION

The recognition of the speaker's identity is another potential use of the glottal radar system [6]. Figure 7 illustrates the procedure of a test-dependent speaker recognition system, including the training phase (dashed lines) and the recognition phase (solid lines).

The testing group consists of 10 speakers (9 male and 1 female speakers). The database included three repetitions of a three-digit utterance "*543*" in Mandarin, which are recorded by both microphone and glottal radars simultaneously. Typical feature extraction methods with Hamming window, MFCC (Mel Frequency Cepstral Coefficients) and SVM (Support Vector Machine) were chosen here [7], and the parameters values applied in this experiment are listed in Table I. All the procedures including data collection, end-point detection,

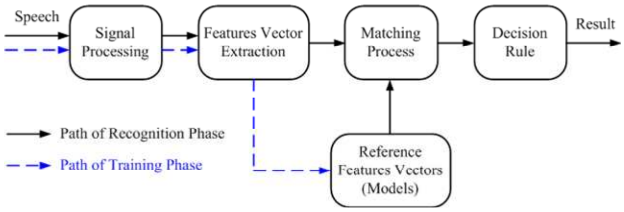

Fig. 7. A typical procedure for speaker recognition.

TABLE I
PARAMETER SETTING OF THE SPEAKER RECOGNITION

Pre-emphasis	α=0.95
Windowing	256-point Hamming window
Frame number	40
Features vector extraction	12-dimensional MFCCs
Training and classification model	SVM
Training/Identification speech	"543" in Mandarin
Number of speaker	10
Number for training per speaker	10
Number for identification per speaker	5

MFCC features vector extraction, SVM classification are all accomplished by Matlab software.

In order to investigate the immunity of glottal radar signal against background acoustic noise, the experiments were conducted in two environments: noiseless and moderate noisy (15-dB SNR). The experimental results for speaker recognition are shown in Table II, where the identification rate is defined as the ratio of the number of speakers identified to the total number of speakers tested. It reveals that a 100% identification rate can be achieved for both acoustic signal and glottal signal in noiseless environment. When the background noise with 15-dB SNR is introduced, the identification rate from acoustic data is dropped to 82%, while the glottal signal still remains 94%.

It should be mentioned that the identification rate from the glottal data is slightly degraded in the later case, which is believed due to the displacement from handhold antenna or body movement. As we mentioned previously in Sec. II, this glottal radar system is sensitive to how and where it attaches to a human subject. With a more solid holding fixture, the antenna displacement problem can be alleviated.

V. CONCLUSION

In this paper, we have proposed two potential applications based on the developed glottal radar system. The experimental results reveal the great potential of the wireless glottal radar signal in high background noise environment where conventional acoustic signal is inacceptable. Therefore it can be applied in mobile handset for communication quality enhancement or the speaker recognition for security system. Further investigations on other applications, such as medical use, are still ongoing.

REFERENCES

[1] J. F. Holzrichter, G. C. Burnett, L. C. Ng, and W. A. Lea, “Speech Articulator Measurements Using Low Power EM Wave Sensor,” *Journal Acoustic Society America*, vol. 103, no.1, pp. 622-625, Jan. 1998.

[2] G. C. Burnett, “The physiological basis of glottal electromagnetic micropower sensor (GEMS) and their use in defining an excitation function for the human vocal tract”, Ph.D. dissertation, University of California, Davis, 1999.

[3] C. Liang, “Novel microwave and millimeter-wave radar technologies and applications”, Ph.D. dissertation, University of California, Davis, 2001.

[4] Y. Yang, “Radar based speech analysis system”, M.S. thesis, University of California, Davis, 2000

[5] [C.S. Lin, S.F. Chang, C.C. Chang, and C.C. Lin, “Microwave human vocal vibration signal detection based on Doppler Radar Technology", *IEEE Trans. Microw. Theory Tech*., vol. 58, no. 8, pp. 2299-2306, Aug. 2010.

[6] T. J. Gable, “Speaker verification using acoustic and glottal electromagnetic micropower sensor (GEMS) data”, Ph.D. dissertation, University of California, Davis, 2000.

[7] L.R. Rabiner and B.H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, New Jersey, USA, 1993.