INTERSPEECH 2017: Show & Tell Contribution
August 20–24, 2017, Stockholm, Sweden

Low-Frequency Ultrasonic Communication for Speech Broadcasting in Public Transportation

Kwang Myung Jeon, Nam Kyun Kim, Chan Woong Kwak, Jung Min Moon, and Hong Kook Kim
School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Korea
`{kmjeon, skarbs001, chanwoong, jungmin7757, hongkook}@gist.ac.kr`

Abstract

Speech broadcasting via loudspeakers is widely used in public transportation to send broadcast notifications. However, listeners often fail to catch spoken context from speech broadcasts due to excessive environmental noise. We propose an ultrasonic communication method that can be applied to loudspeaker-based speech broadcasting to cope with this issue. In other words, text notifications are modulated and carried over low-frequency ultrasonic waves through loudspeakers to the microphones of each potential listener's mobile device. Then, the received ultrasonic stream is demodulated back into the text and the listener hears the notification context by a text-to-speech engine embedded in each mobile device. Such a transmission system is realized with a 20 kHz carrier frequency because it is inaudible to most listeners but capable of being used in communication between a loudspeaker and microphone. In addition, the performance of the proposed ultrasonic communication method is evaluated by measuring the success rate of transmitted words under various signal-to-noise ratio conditions.

Index Terms: low-frequency ultrasonic communication, speech broadcasting, text-to-speech

1. Introduction

Speech broadcasting is widely used in modes of public transportation such as buses, subways, and airplanes to send notifications to passengers in the form of a spoken message [1]. In general, speech broadcasting is simply conducted by playing a pre-recorded or spoken message through the loudspeakers embedded in the vehicle. However, various types of noise during transportation and/or passengers might degrade the intelligibility of the broadcast notifications for those who need to hear it. Such issues might be overcome by providing hearing devices such as noise canceling headphones for each passenger, but this could be a limited solution because 1) the cost of providing such hearing devices to every passenger is exceedingly high for most cases; 2) hearing devices cannot filter out speech-like noises that some passengers might make. Therefore, another solution is demanded to improve the intelligibility of spoken notifications in noisy environments.

Recently, low-frequency ultrasonic (LFU) communication, which utilizes the 19–22 kHz frequency band, has been applied in various indoor services [2]–[4]. This LFU frequency range is inaudible and can be generated by commercial off-the-shelf (COTS) loudspeakers and microphones [2]. Therefore, no additional hardware or operating system is required because low-frequency ultrasound is utilized in any audio interface–equipped device. Moreover, LFU communication is highly robust to various kinds of noise because the carrier frequency of LFU communication is rarely overlapped by bandwidth of noise signals [2]. Thus, LFU communication has been applied in numerous applications such as near-field data transmission [2], TV-based advertisement systems [3], and acoustic markers for augmented reality (AR) devices [4].

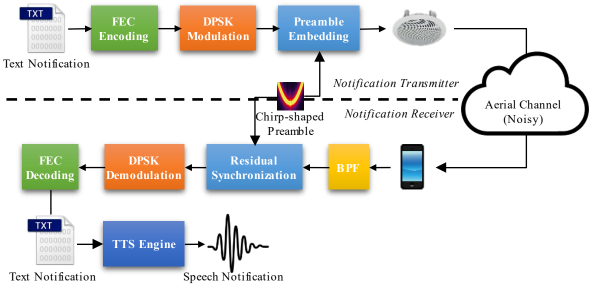

Figure 1: *Procedure of speech broadcasting using the proposed LFU communication.*

This paper proposes an LFU communication method for speech broadcasting in excessively noisy public transportation environments. Figure 1 shows the procedure of speech broadcasting using the LFU communication method. The proposed method first modulates a text notification using a differential phase shift keying (DPSK) scheme with forward error correction (FEC). The modulated text sequence is then transmitted through loudspeakers with a chirp-shaped preamble. The transmitted signals are recorded under the environmental noise by each passenger's smart device microphone. Next, the recorded signals are synchronized with a chirp-shaped preamble template in a residual domain. After synchronization, FEC decoding and DPSK demodulation are performed to restore the text notification. Finally, the text notification is synthesized into a speech signal by a text-to-speech (TTS) engine embedded in the smart device. Finally, the passenger can hear the notification, even in an excessively noisy transportation environment.

2. Notification transceiver

2.1. Notification transmitter

The notification transmitter in the proposed LFU communication method sends a text sequence in a binary format though the LFU wave via three processing steps: FEC encoding, DPSK modulation, and preamble embedding. First of all, the P binary bits to be transmitted at the i-th symbol are grouped as $\mathbf{b}_i = [b_i^1 \cdots b_i^p \cdots b_i^P]$, where P is set to 24 in this paper. Next, $\mathbf{b}_i$ is encoded by a FEC scheme, and additional error correction bits are attached. Specifically, a perfect binary Golay code, G_{23}, is applied as the FEC scheme [5]. In other words, $\mathbf{b}_i$ is reshaped into an $O \times 12$ matrix and is encoded by the (23, 12) Golay block code; thus, the $O \times 23$ encoded block for the i-th symbol,

$\tilde{\mathbf{b}}_i$, is generated. After FEC encoding is completed, each encoded bit of $\tilde{\mathbf{b}}_i$ is modulated by a DPSK scheme. Next, differential encoding is conducted on each $\tilde{b}_i^q$ as in [6]:

$$d_i^q = \tilde{b}_i^q \nabla d_{i-1}^q \tag{1}$$

where d_i^q is the q-th differentially encoded bit at the i-th symbol. Moreover, ∇ in (1) indicates the bitwise exclusive or XOR operation. Next, d_i^q is then modulated in the form of sinusoidal signals by the windowed DPSK $c_i(n)$, given by

$$c_i(n) = w(n)cos\left(2\pi f_c n/f_s + \pi(1 - d_i^q)\right) \tag{2}$$

where n indicates the sample index from 1 to the symbol length, T_b. Moreover, f_c and f_s are the carrier and sampling frequencies, respectively. Here, f_s is set to 44.1 kHz to support the frequency range of most commercial microphones and speakers. Accordingly, f_c can be set any frequency in the rage of 18–22.05 kHz; it is set to 20 kHz in this paper. In addition, $w(n)$ is a squared Hanning window that can smooth discontinuity between subsequent symbols to avoid unwanted impulsive noise at the modulated signal. Note that T_b is set to 810 for reliable communication in excessively noisy environments.

After that, the chirp-shaped preamble, $p(n)$, is designed as a concave form of 166 ms length and with the frequency range of 18–22.05 kHz. Next, the transmission stream is defined as $\mathbf{s} = [\mathbf{p}, \mathbf{c}]$, where $\mathbf{p}$ and $\mathbf{c}$ are the vectors of $p(n)$ and $c_i(n)$ for every n and i. Finally, $\mathbf{s}$ is transformed into analogue signals by a digital-to-analogue converter (DAC), and these signals are then spread into the aerial medium.

2.2. Notification receiver

In the notification receiver, the input signal recorded by the microphone equipped with the smart device of each passenger, $y_i(n)$, is the superposition of $s_i(n)$, where the elements of $\boldsymbol{s}$, and additive environmental noise $z_i(n)$, such as

$$y_i(n) = h(n) * s_i(n) + z_i(n) \tag{3}$$

where $h(n)$ indicates the impulse response of the propagation channel from the LFU transmitter to the receiver. The $s_i(n)$ in (3) can be estimated from $y_i(n)$ by again applying the carrier frequency band-pass filter. That is, $z_i(n)$ in $y_i(n)$ is filtered out because $z_i(n)$ rarely overlaps with $s_i(n)$ at the carrier frequency band [2]. Next, the filtered receiver signal, $\hat{s}_i(n)$, is synchronized with the preamble template, $p(n)$. Here, the impulsive signal detection in a residual domain, which is proposed in [7], is conducted on the cross correlation between $p(n)$ and $\hat{s}_i(n)$ to determine the lag, L_p, for the synchronization. After synchronization is finished, $\hat{s}_i(n - L_p) = \hat{c}_i(n)$ and demodulation and decoding is conducted on $\hat{c}_i(n)$ to recover the binary sequence of the text notification, $\mathbf{b}_i$. Finally, the recovered text notification is synthesized into speech by a TTS engine implemented in the mobile device, which provides spoken notification to each passenger.

3. Performance evaluation

The performance of the proposed method was evaluated by measuring the word error rate (WER) in the transmitted text notification under various SNR conditions. In the evaluation, airplane noise was artificially mixed with the LFU transmission signal, where the average SNR of the LFU frequency band (LFU-SNR) as set to -10, -5, 0, 5, and 10 dB. Note that the LFU transmission signal was generated by modulating a 100-word sequence and the WER under each LFU-SNR condition was evaluated by repeating the transmission 10 times; thus, the total number of transmitted words was 1000 for each LFU-SNR.

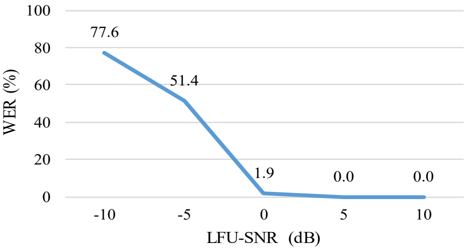

Figure 2: *WER of the notification transceiver*·

Figure 2 shows the WER of the proposed LFU-based notification transceiver according to different LFU-SNRs. As shown in the figure, the proposed method could transmit text notification at a 0 dB LFU-SNR with average WER of 1.9%.

4. Demonstration and conclusion

We designed a demonstration system for the proposed speech broadcasting method that consisted of a notification transmitter with channel simulation functionality and notification receiver. The notification transmitter was developed to have the graphical user interface (GUI) so that it could operate on a low-end laptop with a commercial loudspeaker. Moreover, the notification receiver was developed with the Android platform to run on mid-end smartphones with embedded microphones. The demonstration system is planned to be introduced in a Show and Tell session.

5. Acknowledgements

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the government of Korea (MSIP) (No. 2015R1A2A1A05001687).

6. References

[1] E. Bekiaris, M. Panou, E. Gaitanidou, A. Mourouzis, and B. Ringbauer, “Intelligent mobility and transportation for all,” *The Universal Access Handbook*, CRC Press, pp. 1–29, 2009.

[2] K. M. Jeon, H. K. Kim, and M. J. Lee, “Non-coherent low-frequency ultrasonic communication system with optimum symbol length,” *Int. J. Distrib. Sensors Netw.*, vol. 2016, Article ID 9713180, May 2016.

[3] H. Lee, T. H. Kim, J. W. Choi, and S. Choi, “Chirp signal-based aerial acoustic communication for smart devices,” in *Proc. of IEEE Conf. Comput. Commun. (INFOCOM)*, Hong Kong, China, pp. 2407–2415, Apr. 2006.

[4] K. M. Jeon, C. J. Chun, H. K. Kim, and M. J. Lee, “Application of low-frequency ultrasonic communication to audio marker for augmented reality,” in *Proc. of IEEE Conf. Consum. Electron. (ICCE)*, Las Vegas, LV, pp. 130–131, Jan. 2017.

[5] M. Elia. “Algebraic decoding of the (23,12,7) Golay code,” *IEEE Trans. Inf. Theory*, vol. 33, no. 1, pp. 150–151, Jan. 1987.

[6] B. Sklar, *Digital Communications: Fundamentals and Applications*. vol. 2. NJ: Prentice Hall, Jan. 2001.

[7] K. M. Jeon, D. Y. Lee, H. K. Kim, and M. J. Lee, “Acoustic surveillance of hazardous situations using nonnegative matrix factorization and hidden Markov model,” in *Proc. Audio Engineering Society 137th Convention*, Berlin, Germany, Preprint 9203, Oct. 2014.