Technology Reports

Acoustic Communication System Using Mobile Terminal Microphones

Hosei Matsuoka, Yusuke Nakashima and Takeshi Yoshimura

DoCoMo has developed a data transmission technology called "Acoustic OFDM" that embeds information in speech or music and transmits those sound waves from a loudspeaker to a microphone. It modifies an OFDM signal and superposes on speech or music without significantly degrading the quality of original sound. This technology dramatically improves the amount of information that can be transmitted compared to existing audio-watermarking technology.

1. Introduction

The spread of two-dimensional (2D) codes has made it possible to use a mobile terminal's built-in camera to read in a wide variety of information such as Uniform Resource Locators (URLs) and telephone numbers. In particular, 2D codes that include URLs to Web sites containing reports, documents, and other material have made it easy to access information on the Web from a mobile terminal. Consequently, if the information stored in 2D codes could be embedded in the speech and music of television and radio broadcasts and picked up by the microphone of a mobile terminal, users would have even more occasions for accessing information from their mobile terminals.

However, existing techniques for transmitting information over sound waves based on audio-watermarking technology[*1] can only transmit about one character per second. At this speed, several tens of seconds would be needed to transmit even a simple URL far exceeding a response time comfortable for the user. And while techniques using ultrasonic waves can transmit information at high speeds, there is little audio equipment on the market that can record and play back ultrasonic waves. This and the fact that signals in the ultrasonic band cannot be transmitted

*1 Audio watermarking: A technology for embedding information in speech and music in a way that cannot be perceived by the human ear.

in television and radio broadcasts makes for practically no situations in which ultrasonic waves can be used.

Against the above background, we set out to develop acoustic data transmission technology satisfying the following conditions.

- A URL or simple text information can be transmitted in 1 to 2 seconds.
- Information can be transmitted by sound waves in the audible band[*2] using general commercial loudspeakers and mobile-terminal microphones.
- Transmit signals can be superposed on speech and music without discomforting the human ear.

In this article, we first explain the problems in existing acoustic data transmission technology. We then describe a new acoustic data transmission technology called Acoustic Orthogonal Frequency Division Multiplexing (OFDM) and present the results of a transmission experiment. We also outline a prototype system using Acoustic OFDM and touch upon future developments.

2. Acoustic Data Transmission Technology

There are three main types of audio-watermarking technology used for copyright protection of digital content: echo hiding, spread spectrum and frequency patchwork. With these techniques, the following problems arise in relation to the propagation of information through air.

1) Echo Hiding

Using the characteristic that the human ear cannot perceive short echoes, this technique identifies transmitted bits by altering echo delay and amplitude [1]. In aerial propagation, however, a variety of echoes can occur due to reflected waves and the damped vibration of loudspeakers making this technique inapplicable.

2) Spread Spectrum

Using a psychoacoustic model[*3] [2], this technique computes a frequency-masking[*4] threshold value, multiplies the transmit signal by a Pseudo-random-Noise (PN) series[*5], and superposes the signal spread across the entire frequency band so as to fall below the frequency-masking threshold value [3]. This technique is quite robust with respect to aerial propagation and environmental noise, but since the signal cannot be extracted unless the spreading factor of the PN series is made high, transmission speed inevitably drops (to about 10 bit/s).

3) Frequency Patchwork

This technique superposes the transmit signal by choosing any two frequency bands and then increasing the power of one and attenuating the other so as to create statistical bias [4]. It is capable of information-transfer rates of 40 bit/s, but anything higher can significantly degrade sound quality.

At the same time, using ultrasonic waves to transmit information means that effects on human hearing need not be considered and that transmission speeds faster by an order of magnitude can be achieved since the available frequency band is broader than the audible band. But as explained above, the application of ultrasonic waves requires the use of special playback and recording devices, which would raise the cost of system adoption and diffusion. From this point of view, the adoption of ultrasonic waves is essentially unrealistic. The new acoustic data transmission technology proposed below (Acoustic OFDM) can achieve a practical, working system from the viewpoints of transmission speed, effects on the human ear, adoption cost, and ease of diffusion.

3. Acoustic OFDM

Acoustic OFDM applies orthogonal frequency division multiplexing, a promising transmission scheme for next-generation mobile communications. The OFDM system transmits multiple narrow-band signals in parallel through frequency multiplexing achieving excellent spectrum efficiency[*6]. It can easily cope with the interfering effects of delayed waves making it an effective system for acoustic communications robust to reflected waves.

A key feature of Acoustic OFDM is the transformation of a noisy OFDM modulated signal into an unobtrusive sound and the superposition of that signal on speech or music. In general, sound corresponding to a flat power spectrum often sounds like noise creating an unpleasant feeling in the ear. On the other hand, sound corresponding to power that is biased to particular

*2 Audible band: The range of acoustic frequencies that a human being with normal sense of hearing can hear. Generally, from 20 Hz to 20 kHz.
*3 Psychoacoustic model: Human hearing characteristics modeling aural sensitivity, masking effects, etc.
*4 Frequency masking: Using the effect whereby the sound of neighboring frequencies can create a disturbance, this technique prevents the perception of quiet sounds at frequencies near those of loud sounds.
*5 PN series: The bit string constituting pseudo-random noise. Because PN periodically repeats a previously determined bit string, a PN series can be easily self-synchronized.
*6 Spectrum efficiency: The number of data bits that can be transmitted per unit time and unit frequency band.

frequencies sounds more like a tone than a noise. In a normal OFDM modulated signal, power is uniform across all subcarriers[*7] resulting in a noise-like sound. This unpleasant noise can, however, be converted to a tone in harmony with the speech or music by combining the signal with the speech or music power spectrum so as to superpose the power of each subcarrier.

Figure 1 shows the basic modulation method of Acoustic OFDM. First, the original sound source (Fig. 1 (1)) is subjected to a Fourier transform[*8] to determine the frequency spectrum, and high-frequency components are removed by a Low Pass Filter (LPF) to generate a low-frequency audio signal (2). Next, an OFDM modulated signal (3) is generated by modulating high-frequency subcarriers by the transmit signal. This OFDM modulated signal is then combined with the spectral envelope of the original sound source to adjust the power of the subcarriers and generate a high-frequency audio signal (4). Finally, the low-frequency audio signal and high-frequency audio signal are combined to generate a synthesized audio signal (5), and this signal is output from a loudspeaker.

The audible band for human beings is generally said to extend up to 20 kHz, but the frequency characteristics supported by commercially available loudspeakers and mobile-terminal microphones do not normally extend that high. There are many microphones, for example, that can only pick up sound to about 10 kHz. In addition, the frequencies allocated to the low-frequency signals of an original sound source must be greater than 4 kHz to maintain good sound quality, while the frequencies allocated to OFDM modulated signals range from 5 to 10 kHz at most. However, a frequency bandwidth of 5 kHz is considered to be capable of an information transfer rate greater than 1 kbit/s even when taking an OFDM guard interval[*9] and error correction signal into account. A URL or simple text data could therefore be transmitted in 1 to 2 seconds.

This basic transmission system is combined with techniques for solving problems associated with sound waves and techniques that take effects on hearing into account. The following describes these techniques in detail.

3.1 Frame Boundary

If using phase modulation like Binary Phase Shift Keying (BPSK)[*10] and Quadrature Phase Shift Keying (QPSK)[*11] to modulate each OFDM subcarrier, a phase discontinuity will occur between OFDM frames creating a sound offensive to the ear. This phase discontinuity between OFDM frames must therefore be mitigated. A typical OFDM frame consists of a data-signal section and a preceding guard interval, the latter of

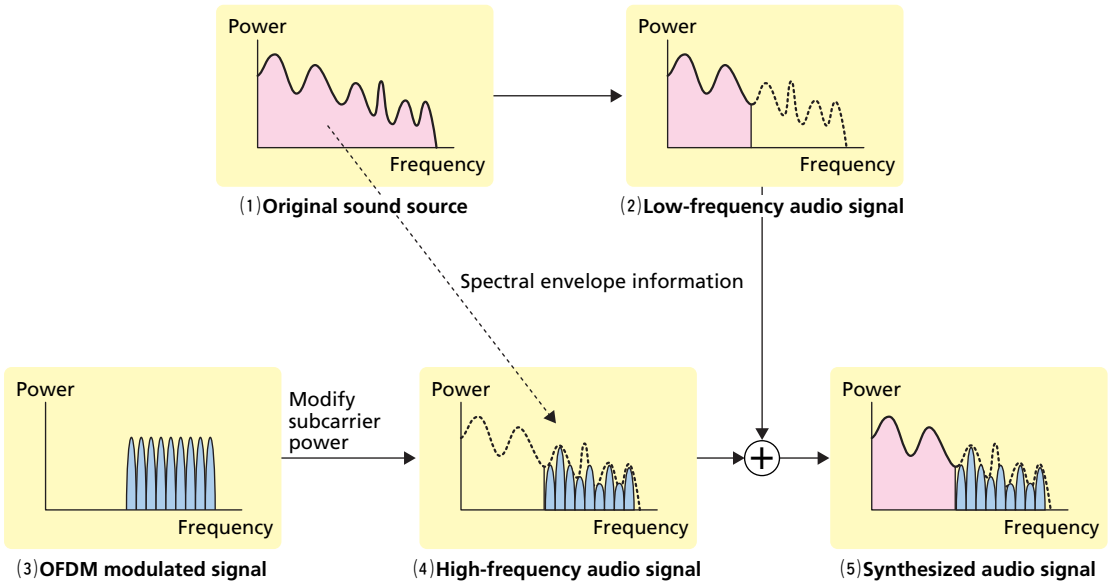

Figure 1 Acoustic OFDM modulation method

*7 Subcarrier: Each carrier in a multi-carrier modulation system that transmits bits of information in parallel over multiple carriers.
*8 Fourier transform: A process that extracts the frequency components making up a signal and their respective ratios.
*9 Guard interval: An interval of fixed duration inserted between symbols to prevent interference between symbols (see *16) caused by delayed waves.
*10 BPSK: A digital modulation method that allows transmission of 1 bit of information at the same time by assigning one value to each of two phases.
*11 QPSK: A digital modulation method that allows transmission of 2 bits of information at the same time by assigning one value to each of four phases.

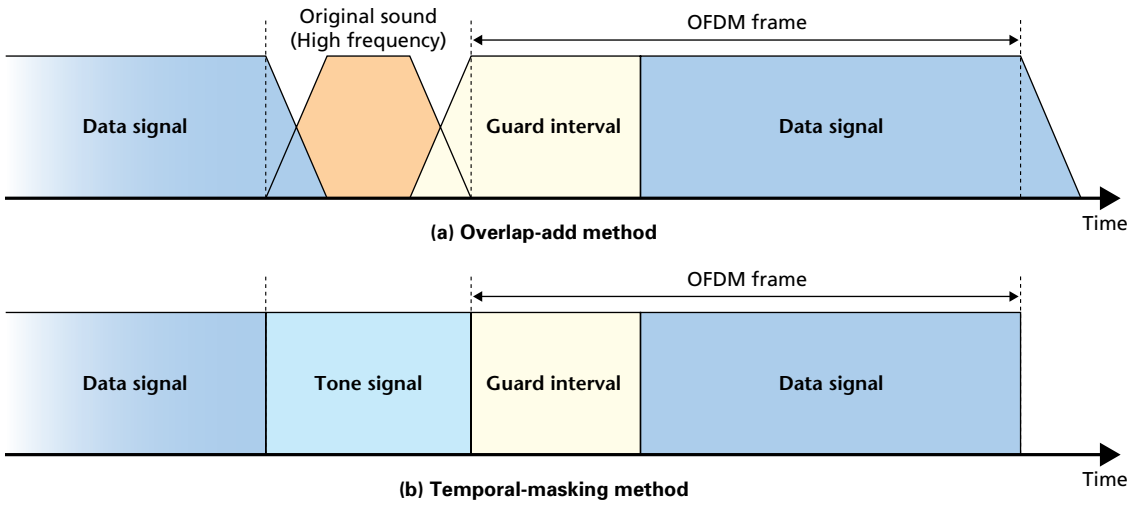

Figure 2 Acoustic OFDM signal

which is formed by copying the back part of the data-signal section. There is also an interval that is inserted at each frame boundary to provide a smooth signal connection. One effective means of inserting this interval is the overlap-add method that overlaps the frame with a trapezoidal window[*12] of the original high-frequency signal (**Figure 2** (a)). Alternatively, temporal masking[*13] can be used to insert a tone signal at the frame boundary to make the grating noise at the discontinuous section difficult to hear (Fig. 2 (b)). In this case, a tone signal would be heard by the user instead of an offending noise. It is possible to create a melody here by selecting appropriate tone-signal frequencies at each frame boundary. The melody can then be used as a signal to the user that a data signal is included in the audio signal.

3.2 Frame Synchronization

Demodulating an OFDM signal at the receiver requires the detection of OFDM frame boundaries. One method for doing this applies the correlation between the guard interval and OFDM modulated signal, but here, accuracy will drop if delayed waves caused by reflection and other factors are present. For this reason, a frame-synchronization signal is added. Making use of a psychoacoustic model [5], this signal is superposed on the low-frequency signals of the speech or music below the frequency-masking threshold. **Figure 3** shows the

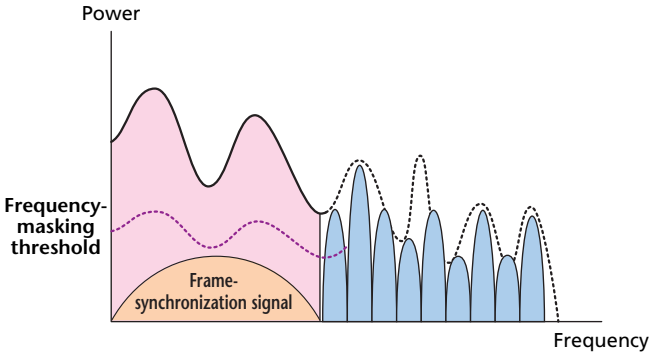

Figure 3 Spectrum of frame-synchronization signal in Acoustic OFDM

spectrum of the frame-synchronization signal in Acoustic OFDM. This process begins by computing the frequency-masking threshold of speech or music low-frequency signals. It then adjusts a frame-synchronization signal, which has been spread across the low-frequency band by a PN series, to a level below the frequency-masking threshold and finally superposes it on the audio signal. The sound associated with the frame-synchronization signal is consequently imperceptible to the human ear. The receiver can now compute the correlation between the received signal and the PN series. The point at which correlation is highest is taken to be the beginning of the OFDM frame enabling demodulation to be performed.

*12 Trapezoidal window: A window function that attenuates both sides of a signal interval to smooth out a signal divided at fixed intervals.
*13 Temporal masking: The effect whereby sound before and after a loud sound becomes difficult to hear; it extends about 5 to 20 ms before and about 50 to 200 ms after.

3.3 Stereo Playback

Devices for playing back speech or music are often equipped with two loudspeakers (left and right) for stereo playback. Here, to play back a monaural signal, the same signal would be played back through both loudspeakers, and to play back a stereo signal, the "left" signal and "right" signal would be played back through respective loudspeakers. In this regard, the simply playback of an Acoustic OFDM transmit signal through two loudspeakers would produce noticeable frequency-selective fading[*14] due to multipath interference. Accordingly, to accommodate stereo playback, the transmit signal to be played back through the left and right loudspeakers should be generated as a stereo signal. At the same time, there is usually only one microphone installed on a mobile terminal meaning that any audio picked up will be a monaural signal. A transmit diversity scheme for two loudspeakers and one microphone should therefore be effective here.

Transmit diversity applies Space Time Block Coding (STBC)[*15] [6] with a coding rate of 1 as shown by matrix G in equation (1).

$$G = \begin{pmatrix} s_1 & s_2 \\ -s_2^{*} & s_1^{*} \end{pmatrix} \tag{1}$$

Here, s^{*} denotes the complex conjugate of s, and s_1 and s_2 denote transmit symbols[*16] at times nT and (n+1)T (T: frame length), respectively. Loudspeaker-L transmits s_1 and $-s_2^{*}$ and loudspeaker-R s_2 and s_1^{*} at those times, respectively. At the receiver, the signal is detected by equations (2)-(5) using receive symbols r_1 and r_2 at times nT and (n+1)T and transfer functions h_L and h_R received from loudspeakers L and R.

(Received signal)

$$r_1 = h_L s_1 + h_R s_2 \tag{2}$$

$$r_2 = h_R s_1^{*} - h_L s_2^{*} \tag{3}$$

(Transmit diversity decoding)

$$\hat{s}_1 = h_L^{*} r_1 + h_R r_2^{*}$$
$$= h_L^{*} (h_L s_1 + h_R s_2) + h_R (h_R^{*} s_1 - h_L^{*} s_2) = (|h_L|^2 + |h_R|^2)\, s_1 \tag{4}$$

$$\hat{s}_2 = h_R^{*} r_1 - h_L r_2^{*}$$
$$= h_R^{*} (h_L s_1 + h_R s_2) - h_L (h_R^{*} s_1 - h_L^{*} s_2) = (|h_L|^2 + |h_R|^2)\, s_2 \tag{5}$$

In the above, $\hat{s}$ denotes the symbols separated by transmit-diversity decoding. With these symbols, s_1 and s_2 can be detected and a transmit-diversity effect obtained. When transmitting the same signal from two loudspeakers, there are many points at which the received signal cannot be extracted due to interference. But when using transmit diversity, receive-signal power can be heightened at any point. Furthermore, considering that the directivity of high-frequency sound waves is sharp, stereo playback by transmit diversity can broaden the range of the transmit signal.

3.4 Frequency Offset and Doppler Shift

In acoustic communications, as in radio communications, deviation in clock frequencies between the transmitter and receiver gives rise to frequency offset[*17]. In fact, deviation in clock frequency is even greater for acoustic equipment since audio playback and recording devices are not originally designed for communications. There are cases in which a frequency offset of 5,000 parts per million (ppm)[*18] is generated between devices. Also, when performing OFDM modulation and demodulation at frequencies in the audible band, offset frequencies can be noticeably different at each subcarrier. For example, given a 5,000-ppm clock-frequency offset between transmitter and receiver, an offset of 25 Hz can occur for a 5-kHz subcarrier and one of 50 Hz for a 10-kHz subcarrier. As a result, the method generally used for correcting frequency offset in radio communications by multiplying a fixed-frequency sinusoidal wave cannot be applied. To correct for frequency offset here, pitch conversion[*19] based on resampling is needed. In addition, the Doppler shift[*20] generated by fluctuation in the location of a pickup microphone cannot be ignored. Given a sonic speed of about 340 m/s, which is about one-millionth the speed of radio waves, a Doppler shift is easily noticeable for even a slight fluctuation in location. The following describes a resampling method for simultaneously correcting this Doppler shift

*14 Frequency-selective fading: A phenomenon in which signal power drops at fixed frequency periods due to interference caused by reflected waves or other delay waves.

*15 STBC: An encoding scheme using transmit-diversity technology. It can separate spatially multiplexed signals through temporal and spatial correlation.

*16 Symbol: In this article, the smallest unit of transmitted data. In QPSK, for example, there are 2 bits of information per symbol.

*17 Frequency offset: In this article, shift in frequency of carrier due to deviation in oscillator clock frequency between the transmitter and receiver.

and the above frequency offset.

First, a pilot signal to be used as a reference for correcting Doppler shift is set on the transmit side at a frequency higher than the frequency band of the OFDM modulated signal. This frequency, denoted as f_i, is known on the receive side. Now, on the receive side, the pilot signal is extracted from the received audio signal through a High Pass Filter (HPF) and FM-demodulated at frequency f_i to detect temporal fluctuation in frequency.

Given that the mobile terminal is moving at a velocity $v(t)$ with respect to a loudspeaker, frequency f_o of the pilot signal detected at the mobile terminal can be given by equation (6) due to the Doppler effect.

$$f_o = \frac{V - v(t)}{V} f_i = f_i - \frac{v(t)}{V} f_i \text{ [Hz]} \quad (V \text{ is speed of sound}) \qquad (6)$$

If the pilot signal observed at this frequency f_o is FM-demodulated at frequency f_i, angular-frequency shift $z(t)$ at time t can be detected.

$$\text{(FM demodulation)} \qquad z(t) = -2\pi \frac{v(t)}{V} f_i \text{ [radian]} \qquad (7)$$

The Doppler shift can now be corrected by resampling the received signal and performing pitch conversion using this $z(t)$ function. Here, the sampling point for resampling can be calculated from $z(t)$ and sampling frequency f_s using equation (8).

$$\text{(sampling-point shift)} = \frac{z(t) f_s}{2\pi f_i} \text{ [sample}^{*21}\text{]} \qquad (8)$$

Resampling based in this sampling-point shift enables both the Doppler shift and frequency offset described above to be corrected simultaneously. Frequency offset appears as a direct-current component of $z(t)$ in FM demodulation.

4. Acoustic Data Transmission Experiment

The following presents the results of an Acoustic OFDM experiment using frequencies in the 5 to 10 kHz range as a basic Acoustic OFDM transmission system. **Table 1** shows the basic specifications of the loudspeaker and microphone used in the experiment and **Table 2** shows transmission parameters.

Table 1 Specifications of loudspeaker and microphone used in experiment

	Loudspeaker	Microphone
Frequency characteristics	20-20,000 Hz	50-16,000 Hz
Aperture	6.5 cm	3.6 cm

Table 2 Transmission parameters

Sampling frequency	44.1 kHz
No. of quantized bits	16 bit
Subcarrier modulation method	BPSK/QPSK
Subcarrier interval	21.5 Hz
No. of subcarriers	234 + 14 (frequency pilot)
Guard interval	11.6 ms
Playback sound pressure	about 70 dBSPL

Spectrum efficiency at these parameters is 1.25 bit/s/Hz for QPSK and 0.63 bit/s/Hz for BPSK. Multiplying a coding rate for error correction code to the above figures gives the amount of information that can actually be transmitted.

4.1 Propagation Distance

Figure 4 shows the relation between propagation distance and Bit Error Rate (BER). The results shown were obtained by measuring BER when modulating each subcarrier in QPSK and BPSK at propagation distances of 1 to 4 m. Assuming that error correction code can correct errors up to a BER of about 5%, allowable propagation distance would be 3 m for QPSK and 4 m for BPSK.

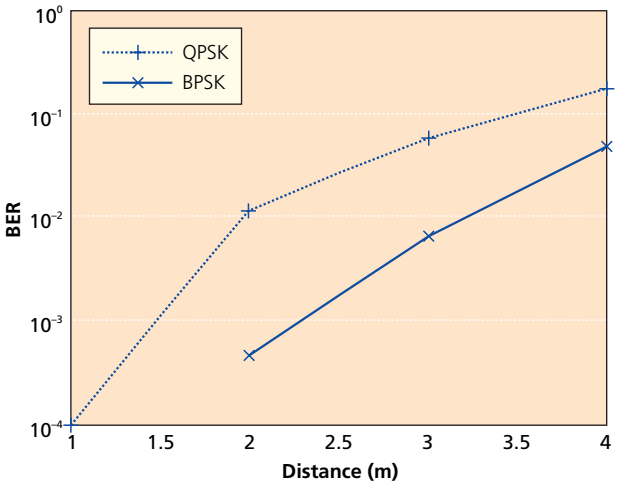

Figure 4 Propagation distance versus BER

*18 ppm: A unit indicating how many millionth of something is present. Although commonly used to represent concentration, it can also be used to indicate the ratio of frequency offset in a carrier.

*19 Pitch conversion: Changing pitch (frequency) by changing the playback speed of the audio signal.

*20 Doppler shift: Shift in frequency of carrier due to the Doppler effect.

*21 Sample: Data obtained in a specific time interval when digitizing. In this article, frame length is expressed as number of samples.

4.2 Directional Angle

The directivity of sound waves becomes sharper as frequency increases and as loudspeaker aperture becomes larger. In Acoustic OFDM, signals are transmitted in a high-frequency band making for sharp directivity. **Figure 5** shows the relation between BER and the angle of arriving sound waves. These results were obtained by measuring BER at directions of 0 to 60° at a propagation distance of 2 m. The aperture of the loudspeaker used in the experiment was 6.5 cm. Assuming here as well that error correction can be performed up to a BER of about 5%, allowable propagation range would be up to 20° for QPSK and 50° for BPSK.

4.3 Frequency Response

Figure 6 shows amplitude and phase characteristics in subcarriers. Examining the amplitude characteristics, a difference of about ±5 dB can be seen due to the effects of loudspeaker/microphone amplitude characteristics and frequency-selective fading. As for phase characteristics, it can be seen that phase deviation differs for each frequency, but since the main reason for this is delayed waves due, for example, to reflection, phase characteristics turn out to be nearly linear (group delay is constant). As a result, phase deviation can be corrected by inserting pilot signals at a certain frequency interval.

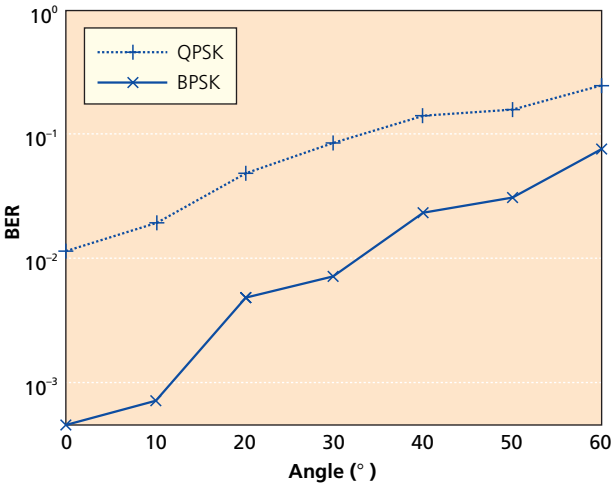

Figure 5 Angle of arriving waves versus BER

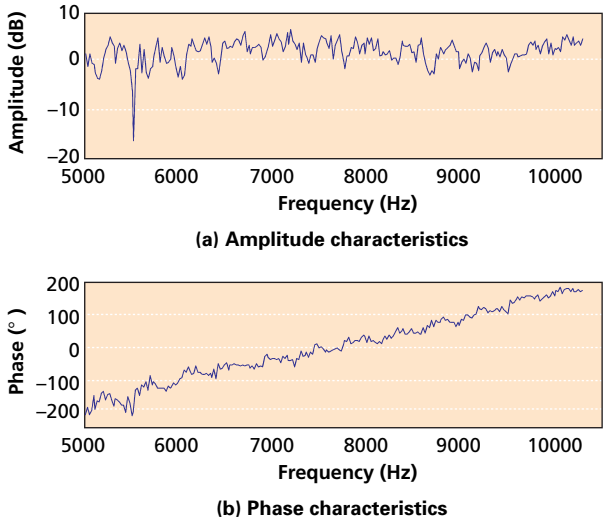

Figure 6 Amplitude and phase characteristics

5. Outline of Prototype System

We implemented a prototype system using Acoustic OFDM. **Figure 7** shows the system configuration. This system makes use of carousel transmission[*22] while embedding simple text information like URLs in speech or music. Here, tone signals can be inserted at OFDM frame boundaries during data transmission using the time masking method described in section 3.1. Selecting appropriate frequencies for these tone signals can produce a melody indicating to the user that data is included in the audio signal. The user can then extract the embedded data by picking up the sound in question for about 1.5 s during any interval in which this signal can be heard.

On the transmitting side, the system inputs a speech or music signal plus the data signal such as a URL to be superposed on the former, and encodes the data signal with Bose Chaudhuri Hocquenghem (BCH) code. This enables the receiving side to perform error correction and extract the original data signal provided that bit error is about 5%. After BCH coding, the system converts the bit string from serial to parallel. The system also subjects the speech or music signal to a Fourier transform to compute the frequency spectrum, and cuts out the high-frequency signals by a LPF. Now, using the frequency spectrum so computed, the system adjusts the power of OFDM subcarriers, performs OFDM modulation on the parallel-con-

*22 Carousel transmission: Repeated transmission of the same data. Once transmission of certain data has been completed, the same data is transmitted again from the beginning.

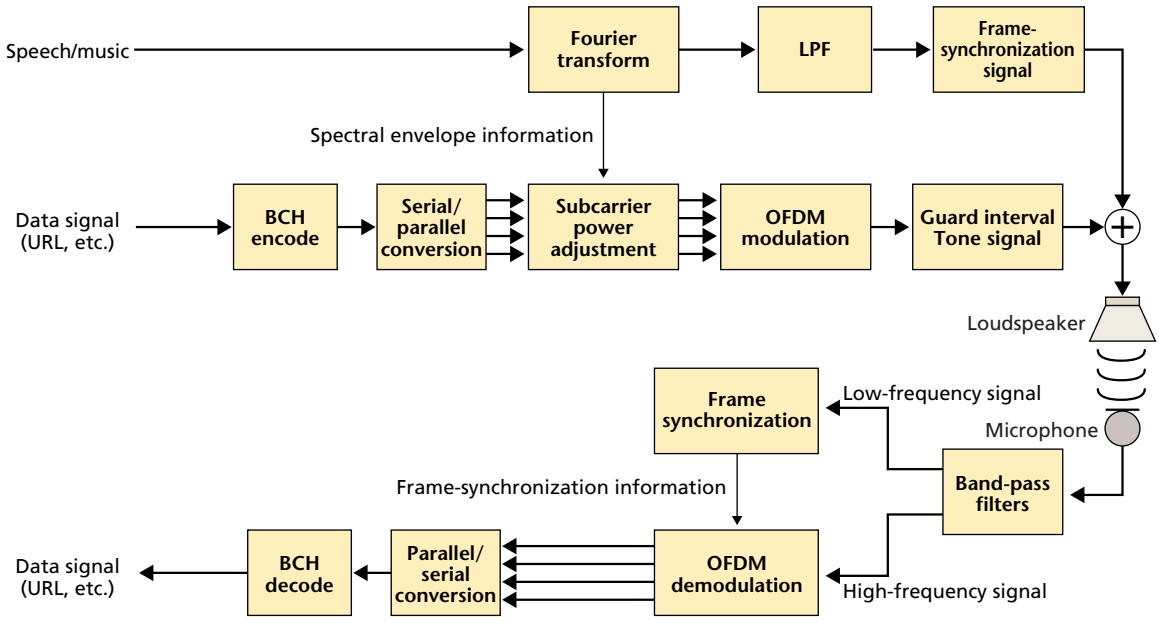

Figure 7 System configuration

verted data signal using these subcarriers, and inserts a guard interval and tone signal. Finally, the system adds a frame-synchronization signal based on a PN series having high self correlation to the low-frequency portion of the speech or music signal at intervals corresponding to the OFDM frame period, and combines the result with the OFDM modulated signal outputting the synthesized signal from the loudspeaker. This output signal is now picked up by the microphone and divided into a low- and high-frequency signal by a band-pass filter. The correlation between the low-frequency signal and the PN series is computed and frame synchronization is achieved by treating the point with the highest correlation as the beginning of an OFDM frame. This process enables the guard interval and the tone signal to be removed from the high-frequency signal for every frame unit and for OFDM demodulation to be performed. The system can now convert the demodulated parallel signals into a serial signal and extract the data in question by BCH decoding. Note here that, regardless of what frame reception begins at, data can be decoded on the receiving side as long as a certain number of frames are received. Since BCH code is a cyclic code, error-correction decoding can be performed even if the bit string of a code word is cycling, and the beginning of a data signal can be detected from the number of cyclic bits in a code word.

Table 3 shows the parameters of the prototype system. We loaded a receive application for a system based on these parameters in a mobile terminal and performed a test to evaluate data acquisition when superposing a data signal on the speech or music of a TV commercial. We found that text information having a length of 72 bytes could be obtained with an identification rate above 90% given a range of 1 m from the loudspeaker.

6. Conclusion

We proposed an acoustic data transmission technology

Table 3 System parameters

Sampling frequency	44.1 kHz
OFDM frame length	2,032 sample
Data length	1,024 sample
Guard interval	600 sample
Tone-signal interval	408 sample
Band-pass filter cutoff frequency	5,512.5 Hz
No. of subcarriers	33 + 4 (frequency pilot)
PN-series length	127
Chip rate	2,756.25 Hz

called "Acoustic OFDM" that satisfies practical requirements from the viewpoints of transmission speed, effects on hearing, and ease of adoption. We presented the results of OFDM transmission experiments using sound waves and outlined a prototype system. In future research, we plan to work on schemes for improving transmission performance and reducing computational volume.

REFERENCES

[1] D. Gruhl, A. Lu and W. Bender: "Echo Hiding," Information Hiding 1996, pp. 295–315.

[2] "Coding of moving pictures and associated audio for digital storage media at up to about 15Mbit/s," ISO/IEC 11172,1993.

[3] L. Boney, A. H. Tewfik and K. N. Hamdy: "Digital watermarks for audio signals," IEEE Ind. Conf on Multimedia Computing and Systems, pp. 473–480, Mar. 1996.

[4] A. Nakayama and S. Iwaki: "HyperAudio: A Man-Machine Interface Technique via the Medium of Sound," No. 088, 2000 (In Japanese).

[5] J. Johnston: "Transform Coding of Audio Signals Using Perceptual Noise Criteria," IEEE Journal on Selected Areas in Communications, Vol. 6, pp. 314–323, Feb. 1988.

[6] S. M. Alamouti: "Simple Transmit Diversity Technique for Wireless Communications," IEEE Journal on Selected Areas in Communications, Vol. 16, pp. 1451–1458. Oct. 1998.