Mobile Maestro: Enabling Immersive Multi-Speaker Audio Applications on Commodity Mobile Devices

Hyosu Kim[†], SangJeong Lee[‡], Jung-Woo Choi[¶], Hwidong Bae[†], Jiyeon Lee[†], Junehwa Song[†], Insik Shin[*]
[†] [*]Dept. of Computer Science, [¶]Dept. of Mechanical Engineering, KAIST, South Korea
[‡]Software Center, Samsung Electronics, South Korea
[†] [¶] [*]{hyosu.kim, khepera, spilist, jy.lee, junesong, insik.shin}@kaist.ac.kr, [‡]sj94.lee@samsung.com

ABSTRACT
The goal of this work is to provide an abstraction of ideal sound environments to a new emerging class of *Mobile Multi-speaker Audio* (MMA) applications. Typically, it is challenging for MMA applications to implement advanced sound features (e.g., surround sound) accurately in mobile environments, especially due to unknown, irregular loudspeaker configurations. Towards an illusion that MMA applications run over specific loudspeaker configurations (i.e., speaker type, layout), this work proposes AMAC, a new *Adaptive Mobile Audio Coordination* system that senses the acoustic characteristics of mobile environments and controls individual loudspeakers adaptively and accurately. The prototype of AMAC implemented on commodity smartphones shows that it provides the coordination accuracy in sound arrival time in several tens of microseconds and reduces the variance in sound level substantially.

Author Keywords
Audio coordination; acoustic sensing; mobile multi-speaker audio appliation

ACM Classification Keywords
H.5.5. Information Interfaces and Presentation (e.g. HCI): Sound and Music Computing; H.5.1. Information Interfaces and Presentation: Multimedia Information Systems

INTRODUCTION
Accurate sound reproduction - as close to the original sound as possible - creates a rich, engaging user experience. Music fans would like to listen to music on fine-tuned high-end audio systems, embracing the feeling of being at a live concert. The widespread use of mobile devices opens an opportunity to enable a new class of applications, namely *Mobile Multi-speaker Audio* (MMA) applications; they utilize loudspeakers on multiple mobile devices to deliver immersive in-situ sound reproduction even in poor acoustic settings (e.g., outdoors). Imagine a group of friends on a camping trip to a mountain. After dinner, one suggests watching an action movie together. They find no video/audio equipment at the camping site but the loudspeakers on their own smartphones. They have little expertise and experience for audio set-up. They would be happy with a click-away construction of a spectacular surround sound for the movie through a MMA support.

Over several decades, many advanced acoustical techniques have been developed for immersive sound reproduction [33]. Sound spatialization is one of the key acoustical techniques to enrich sound reproduction quality, by creating a feeling of a certain location and movement of sound [2]. To reproduce such sound features, various multichannel systems, from 5.1 channel to USC 10.2 and NHK 22.2, as well as holographic reproduction system [10, 18] have been suggested. The sound systems typically simulate two most important auditory cues in human sound perception: the difference in sound arrival time and the difference in sound pressure level[1] requiring a high precision control over both arrival time (less than 1ms) and level (as accurate as possible).

A common, traditional requirement for enabling the highly-accurate acoustic control on these systems is that loudspeaker configurations should be pre-defined at specific locations with a specific type of loudspeakers. Recently, lots of attempts have been made to overcome this limitation, by introducing the object-oriented scheme in audio codecs [7] and designing sound renderers that can synthesize the key auditory cues with flexible loudspeaker locations [28]. Moreover, many home theater systems [1, 27, 31] calibrate speakers automatically to deal with room acoustic problems, while often employing special measurement tools such as measurement-grade mics.

This work aims to provide an illusion that MMA applications run over conventional loudspeaker configurations (i.e., home theater systems) in mobile environments. This requires to deal with several factors. First, mobile devices can be arbitrarily placed with different distances from a listener, forming an irregular layout. Second, a group of users often have heterogeneous mobile devices, leading to the different acoustic characteristics (i.e., audio output power) of embedded loudspeakers. The sounds traveling along different paths can ex-

[*]A corresponding author.

[1]Sound pressure level, sound level for short, is an acoustic pressure of a sound relative to a reference level (e.g., the threshold of hearing). It is logarithmically measured in decibels (dB).

UbiComp '14, September 13 - 17, 2014, Seattle, WA, USA

ACM 978-1-4503-2968-2/14/09...$15.00.
http://dx.doi.org/10.1145/2632048.2636077

perience asymmetric acoustic effects such as diffraction and reflection, depending on room configuration. Auto calibration techniques can be adopted to address the above factors for high precision sound control, while raising new additional challenges. In mobile situations, acoustic environments can vary dynamically due to the mobility of users and/or devices. This entails to perform sensing and calibration continuously and seamlessly. In addition, commodity mobile devices generally come with little support for high precision sound control. For instance, they typically offer a coarse-grained accuracy in clock synchronization (i.e., several milliseconds) and often yield large and unpredictable delays in audio playback.

Thus, the goal of this work is to provide an abstraction of virtual sound environments that can simplify the development of MMA applications. The key underlying technology to support this new abstraction is the ability to sense the acoustic characteristics of the physical sound environments and coordinate each individual audio player adaptively based on the observation. For sensing, it captures loudspeaker-specific characteristics (e.g., clock deviation, acoustic impedance) and layout/room-specific properties (e.g., propagation delay, sound level attenuation). For coordination, it provides finer-grained and continuous audio stream control (e.g., arrival time synchronization and frequency-specific sound level equalization). Thus, whichever loudspeakers are used, wherever they are placed, those techniques are able to deliver audio signals to a listener at designated arrival times and with intended sound pressure levels as accurate as possible to implement advanced acoustical techniques for mobile environments.

Contribution. To this end, this work proposes AMAC, a new *Adaptive Mobile Audio Coordination* system with the following contributions:

- It presents a design of AMAC to support the virtual acoustic environment abstraction by creating an illusion that MMA applications play audio through a set of homogeneous loudspeakers placed in a regular layout with symmetric acoustic effects.

- It provides a prototype implementation of AMAC, along with evidences that AMAC enriches sound reproduction quality that were previously very difficult on commodity mobile devices. A new API introduced by AMAC enables MMA applications to maintain the coordination accuracy in sound arrival time in the order of tens of microsecond and hold a high degree of accuracy in sound level coordination, adapting to the different acoustic characteristics of mobile environments.

- It demonstrates the effectiveness of AMAC with our MMA application prototype, MobileTheater[2], for movie playback with 5.1 surround sound. Our experience with real users is positive. Users report that AMAC helps to create much more immersive experience for music listening and movie watching.

[2]See http://cps.kaist.ac.kr/mobile_maestro for a video illustrating a movie playing back with the surround sound by MobileTheater.

We anticipate that with the multi-speaker audio API we provide in this paper, many more MMA applications can be easily developed.

BACKGROUND AND CHALLENGES

In this section, we describe an emerging class of mobile multi-speaker audio (MMA) applications, outline human sound perception, and present challenges in supporting MMA applications.

Emerging MMA Applications

The widespread use of mobile devices offers an attractive opportunity to enrich sound reproduction through a configuration of multiple loudspeakers. Each individual loudspeaker can emit audio signals, collectively achieving various acoustic effects. Building upon this opportunity, we envision a class of MMA applications will continue to evolve delivering a new experience of in-situ listening.

1) *Contemporary.* Recently, some commercial mobile systems (e.g., Samsung GroupPlay [30]) were introduced to play the same song on multiple loudspeakers to simply amplify the sound volume beyond the limit of a single speaker. This can help to enrich in-situ user experience. For example, a group of fans at a noisy stadium can boost their excitement singing together a cheering song accompanied loudly by multiple mobile devices.

2) *Emerging.* Emerging MMA applications can offer advanced sound features, such as multichannel surround sound [17], directional sound [8], and noise cancellation [9, 14]. As an example, for 5.1 surround sound, each of six mobile devices can play their own audio channel to jointly deliver proper sound spatialization and amplification. Some movie fans can experience an increase in positive emotion when watching their favorite movie scenes at film locations (e.g., watching Hogwarts dining hall scenes in Harry Potter at the University of Oxford) with a richer sound effect.

Understanding Human Auditory Perception

Humans can judge the location of a sound source primarily on the basis of two complementary auditory cues: the interaural level difference (ILD) and interaural time difference (ITD). This psychoacoustic principle, called duplex theory [5], has been intensively studied for the source localization. In particular, various acoustic techniques that can relate ILD/ITDs to sound arrival level difference between multiple loudspeakers have been suggested (e.g., stereophonic law of sine [3] and tangent [4]). The possible variations of ITD and ILD due to the sound source movement depend on the size of the head and frequency, but in most cases, ITD changes within $\pm$1ms range and ILD change occurs within $\pm$20dB scale. In detail, the ILD change of 5dB can occur when the direction of a sound source changes from 0° to 30° at 2.5kHz.

Another important auditory effect appears when multiple correlated sounds arrive within the range of 1ms and 30ms after the first-arriving sound. An auditory event is then dominantly localized by the first-arriving sound and the delayed sounds are perceived as the reverberation of the first sound, when arriving up to 10dB louder than the first sound [21]. The reverberation sounds work positively creating a feeling of natural sound ambience without affecting the sound localization. So, the effect can be used for audio systems to enhance loudness keeping a desired location.

Requirement. As such, humans perceive diverse high-quality acoustic features based on subtle difference in sound arrival time and level. Thus, in order to reproduce the designated immersive sound effect, audio systems require a high degree of accuracy in inter-speaker sound arrival time (up to 1ms) and inter-speaker sound arrival level (as accurate as possible).

Challenges for Enabling MMA Applications

Conventional audio systems recommend specific loudspeaker configurations to meet the requirements for optimum sound quality. In case of the 5.1 channel surround, for example, it is proposed to have six speakers of the same type and keep reference distance from listeners for all speakers [17]. However, such configurations are hardly supported for MMA applications, and this makes it challenging to meet the acoustic requirements in mobile environments due to the following reasons:

- **Different acoustic environments.** In many in-situ environments, users can utilize mobile devices as loudspeakers at different locations, such as indoor or outdoor places, leading to different room acoustics (i.e., diffraction, interference, and reflection) [20]. Those loudspeakers can be placed arbitrarily at different distances from listeners, forming an irregular speaker layout. Since many different users are equipped with different mobile devices, MMA applications will run over heterogeneous mobile devices in most situations. Thus, in mobile environments, the inter-speaker differences in both sound arrival time and sound level are likely to increase due to several factors, including different sound traveling distances, different audio output power, and different sound effects.

- **Dynamically-changing device layouts.** In mobile environments, users and/or devices are expected to move, even while MMA applications are in the middle of audio playback. Such a move can make it necessary to re-configure loudspeakers for better sound reproduction. This leads to sense acoustic environments and coordinate speakers in a continuous and seamless fashion, without distracting users.

- **Little support of commodity mobile devices.** Commodity mobile devices do not offer tight timing guarantees on the operation of playing audio signals, and thus systematically incur large and unpredictable time delays. Such a limited timing support generally causes an unintended sound arrival time difference of high magnitude and high variance between devices.

AMAC SYSTEM DESIGN

The main design goal of AMAC, our adaptive mobile audio coordination system, is to provide an abstraction of ideal sound environments to MMA applications. In other words, AMAC aims to provide the API, on top of which MMA applications deliver rich sound features as if users are listening with fine-tuned audio system, regardless of heterogeneous devices, irregular layouts, or different room acoustics they have. To this end, AMAC is designed to conduct an adaptive, automated, and real-time coordination for immersive sound effects. Importantly, it addresses two main factors: sound arrival time and sound arrival level. The key method for coordinating such factors is to construct a feedback control system using acoustic control signal. AMAC observes what kind of changes are made while the control signal travels through the real-world mobile environments. Based on the observation, AMAC adaptively adjusts the arrival time and the sound level to reproduce designated sound effects.

Overview

The AMAC system consists of two types of entities; one or more audio players, and a single coordinator.

Each player is assumed to have its own audio stream to play through its loudspeaker. For instance, six players can have its own audio signal on each different channel of 5.1 channel surround sound. The key goal of the coordinator is to create a listening sweet spot on a user's location. So, it is placed close to the listener and determines the *playback start time* and *sound level* of each individual player in a way that the sounds emanating from individual players arrive at the coordinator at designated arrival times and with sound levels as accurately as possible no matter which speaker type, which speaker layout, and which room acoustic features.

Towards the goal, the coordination process proceeds as follows: First, all players transmit coherent acoustic control signals to the coordinator to capture acoustic characteristics of mobile environments. Next, in order to adapt to the given environments, the coordinator adjusts the playback start time of each player and constructs a finite impulse response (FIR) filter to tune the level of the sound from each player. Upon receiving the playback start time and the FIR filter from the coordinator, each player begins playing back its own audio stream at the designated time while transforming the stream through the given filter. The transmission and adjustment are processed once at an initial coordination stage, and then, regularly repeated with inaudible acoustic signals during playback to cope with the dynamically-changing environment.

Acoustic control signal

AMAC makes use of acoustic control signals to measure the two types of acoustic responses: *impulse* and *frequency responses*. An impulse response shows the sound transfer process as an energy-time function. On the other hand, a frequency response offers the frequency-specific magnitude of a sound, i.e., how loudly a sound wave arrives at each frequency. We explain later how to determine appropriate playback start time and sound level for each player, based on impulse and frequency response measurements.

MLS control signal. As an acoustic control signal, AMAC employs a maximum length sequence (MLS), which is widely used for acoustic measurements in many different application fields [6]. The MLS signal offers several benefits for MMA scenarios. (1) *Accuracy.* The MLS signal makes it possible to obtain an accurate impulse response even in the presence of noise and reflections. The MLS signal consists of a full range of frequencies with a flat spectrum, and its auto-correlation function is essentially unit-impulse. This allows to obtain an

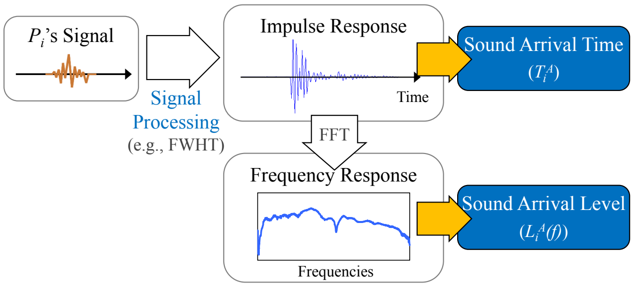


Figure 1: A computation phase of acoustic sound arrival time and level measurement.

accurate impulse response for a full range of frequencies by cross-correlation. (2) *Efficiency.* In general, it is computationally heavy to calculate the impulse response through cross-correlation. However, the MLS signal enables to do it in a more computationally efficient way. Cohn et al. [11] showed the equivalence of the MLS signal to the Walsh-Hadamard transform. This relationship allows an impulse response to be computed by a modified fast Walsh-Hadamard transform (FWHT), lowering a computational complexity from $O(K^2)$ to $O(K \log K)$, where K is the length of a signal.

Sound Arrival Time & Level Measurements

AMAC performs sound arrival time and level measurements in three steps: *preliminary*, *transmission*, and *computation* steps. In the preliminary step, the coordinator and players synchronize their clock times in a coarser-grained level. In the transmission step, players send MLS signals to the coordinator, and in the computation phase, the coordinator calculates impulse and frequency responses from the signals to measure sound arrival times and levels.

Preliminary step. In the preliminary step, AMAC seeks to reduce initial clock time differences between the coordinator and players through a network-based clock synchronization method, called NTP (Network Time Protocol) [23]. In many cases, NTP supports the synchronization accuracy in the order of tens of millisecond [24].

Transmission step. The coordinator requests each player P_i to transmit a MLS signal at time T_i^R at their maximum output power. AMAC seeks to arrange a different time T_i^R for each P_i in order to avoid time-aliasing errors due to the overlap of different MLS signals. There are many reasons for the overlap between the signals. For example, with the coarse-grained clock synchronization in the preliminary step, the actual transmission of P_i can happen some time before or after T_i^R with an error in the order of tens of millisecond. In addition, since sound waves experience reflection, the reverberation of one signal can overlap with another. In most common places for in-situ sound reproduction, the reverberation time would be no longer than 1 second. Considering these, AMAC places two guards intervals, T^{G1} and T^{G2}, before and after T_i^R, to prevent MLS signals from overlapping. Then, AMAC arranges T_i^R for each P_i as follows:

$$T_i^R = T_{i-1}^R + T^{MLS} + T^{G1} + T^{G2},$$

where T^{MLS} is the time length of a MLS signal, T_1^R is equal to the current time + T^{G1}, and T^{G1} and T^{G2} are set to 0.5 and 1 second(s), respectively. Then, the audio signal arriving at the coordinator during $[T_i^R - T^{G1}, T_i^R + T^{MLS} + T^{G2}]$ is considered as the received MLS signal of P_i.

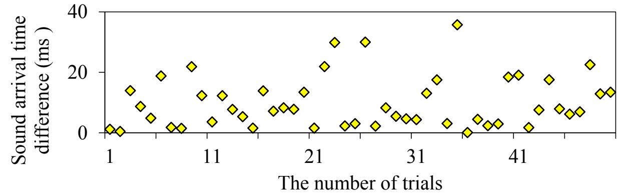


Figure 2: A sound arrival time difference on Android.

Computation step. Upon receiving the MLS signals from all the players, the coordinator computes impulse and frequency responses from each signal (see Figure 1). An impulse response of P_i is derived from the received MLS signal through a signal processing mechanism, i.e., FWHT. Since the response is represented as the energy-time function of sound transmission, it has a maximum peak value when the sound arrives directly (without any reflection) and becomes lower with reverberation. This allows to pinpoint the actual arrival time (T_i^A) of the MLS signal that P_i transmits, as follows:

$$T_i^A = T_i^R - T^{G1} + T_i^{PT},$$

where T_i^{PT} indicates the time instant at which the amplitude of P_i's MLS signal has the peak value. The measured response is also transformed into frequency domain, i.e., a frequency response, by applying the fast Fourier transform to calculate sound level of P_i. As discussed, the frequency response provides the loudness of P_i as a function of frequency. So, AMAC gets a sound arrival level at frequency f, $L_i^A(f)$, from the frequency response.

AMAC SYSTEM IMPLEMENTATION

In this section, we describe the implementation of AMAC on commodity mobile devices, sharing some specific implementation details.

Sound Arrival Time Coordination

As discussed, the difference in sound arrival time is one of the key factors for high-quality sound reproduction. In this subsection, we present the synchronization techniques to enable sound arrival time coordination with a great amount of accuracy on commodity mobile devices.

The propagation delay T_i^D of each player P_i is defined as the time required for a sound to travel from P_i to the coordinator, and it can be estimated as

$$T_i^D = T_i^A - T_i^R.$$

Depending on the speaker layout, room acoustic effects, and audio playing latency, individual players can have different propagation delays to the coordinator, leading to different arrival times even though they emit sounds simultaneously. In order to compensate for different propagation delays, the coordinator determines the playback start time T_i^S of each player P_i as follows

$$T_i^S = T_i^* - T_i^D,$$

where T_i^* is the playback start time for a designated immersive acoustic effect from the coordinator's viewpoint (e.g., 2 seconds after the current time). This way, all the sounds emanating from individual players P_i at T_i^S arrive at the coordinator at T_i^*. In practice, however, it is not easy for P_i to emit its own audio stream exactly at T_i^S because of some technical limitations on commodity mobile devices.

Limitations on Android platform. In Android, the most widely used mobile platform, an API, called *SoundPool*, is used for an audio playback. When *SoundPool* is called, it takes four steps to handle the request; pre-processes the audio stream source, wakes the loudspeaker up, writes pre-processed source to the audio device buffer, and reads the buffer to emit audio stream.

The first problem here is an unpredictable pre-processing time. Since Android tries to pre-process as fast as possible, the latency becomes arbitrarily distributed according to the system's current status (e.g., the degree of resource competition with other applications). Although the latency can sometimes be very low, such randomness is not appropriate for synchronization purpose. Figure 2 shows the sound arrival time difference between two devices running Android. For each device, we first perform clock synchronization with NTP. Devices then emit the MLS signal 50 times by using *SoundPool*, and we measure the sound arrival time differences. Note that the sound arrival times are randomly distributed in a large interval [68μs, 35,760μs].

Second, a large amount of delay is unpredictably introduced to wake up loudspeakers when they are not ready. For energy efficiency, Android forces loudspeakers to sleep when their playbacks finish. The problem is that in AMAC, loudspeakers are required to emit sound twice in a consecutive manner; one for measurement (i.e., the transmission step), and another for actual audio playback. Although loudspeakers will be used right after measurement, they are forced to sleep during the computation step because Android is not aware of the nature of AMAC. Thereby, AMAC needlessly becomes to go through tens of milliseconds to wake the loudspeaker up again.

Last, the device buffer does not immediately read the audio source. The audio device driver checks an empty buffer periodically (e.g., every 5ms on Nexus 4). Because the buffer is empty during the computation step, the audio source written between the checking periods is delayed to play until the next buffer check. Similar to other problems, this makes it difficult to provide a fine-grained control of a playback start time.

A new audio API for a predictable sound reproduce. AMAC solves above problems by adding a new audio playback API called *StableSoundPool*. As explained, the original pre-processing routine takes up to 36ms. Thereby, *StableSoundPool* gives an enough preparation time interval T_i^P (e.g., 100ms) for pre-processing, so that AMAC is able to write the audio stream to the buffer in a predictable manner.

Next, *StableSoundPool* solves the other two problems in a simple, yet effective way. Until the actual playback gets started, *StableSoundPool* writes zeroed data to the audio buffer. It not only enforces loudspeakers to emit a mute sound, but prevents their needless sleep so that the device driver immediately reads the audio stream.

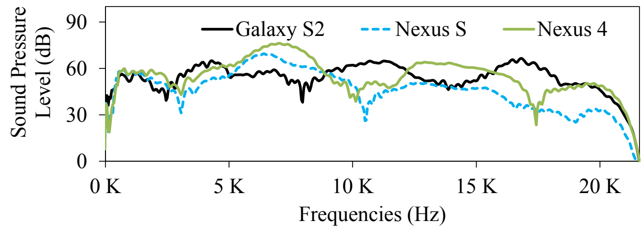


Figure 3: The sound arrival levels of the sounds emanating from different phones but traveling along the same path.

Since the pre-processing time is now stable and additional wake-up delays are removed, AMAC is able to consistently maintain playback start time as $T_i^S + T_i^P$.

Sound Arrival Level Coordination

In addition to the difference in arrival time, another most important factor for immersive sound reproduction is the variance in sound arrival level. In this section, we describe challenges for sound level coordination in mobile environments and present our equalization-based approach to address them.

Challenges. Mobile environments raise many challenges for accurate sound arrival level coordination, such as device heterogeneity and layout irregularity. Heterogeneous devices are typically equipped with their own loudspeakers that have different audio output capabilities, as illustrated in Figure 3. In the figure, three different phones are shown to produce different sound arrival levels when emitting the same audio signal along the same path at their full audio output powers. For instance, Nexus S generates a larger arrival level than Galaxy S2 in the frequency range of [5.4kHz, 9.3kHz], but vice versa in other ranges (i.e., higher than 9.3kHz).

Irregular device layouts yield different distances between the coordinator and players. Since the signal strength decreases inversely proportional to the traveling distance, even homogeneous loudspeakers can produce different sound arrival levels when placed with different distances from a listener. However, the inverse-distance law is not valid in most real situations, in which each sound wave goes through various acoustic effects, such as diffraction and reflection. Therefore, the level coordination should include the extra-compensation of nonidentical frequency responses due to these acoustic effects as well as the device heterogeneity.

Level Equalization. For the sound arrival level coordination, AMAC makes the use of finite impulse response (FIR) filters that enable the frequency-specific sound arrival level control. A coordinator adaptively constructs the FIR filter of each player based on the sound arrival level measurements, and each player applies the filter to its own audio stream. In the acoustic measurement, each player P_i emits the coherent MLS signal at its own maximum output power, and the coordinator measures the sound arrival level $L_i^A(f)$ of every frequency f. Since P_i cannot make more amplified sound arrival level than $L_i^A(f)$ at each frequency f, the coordinator computes a desired additional sound level decay $L_i^D(f)$ for the equalization as

$$L_i^D(f) = \min_{\forall j}[L_j^A(f)] - L_i^A(f).$$

Based on this, sound arrival levels can be equalized without any amplification. However, such $L_i^D(f)$ can cause substantial de-amplification, leading to some significant degradation to user experience. For example, in Figure 3, Nexus S experiences a sudden fall in the sound arrival level at frequency of 11kHz since the loudspeaker of the device emits nearly muted sound at the frequency. Because such hardware-driven limitation could make all devices extremely attenuate the signal strength, we find the deep valleys by using the zero-tracking mechanism and do not select them as the minimum level.

A FIR filter F_i is constructed based on the additional sound level decay $L_i^D(f)$. If each player emits the sound wave applying $L_i^D(f)$ of every frequency f to its audio stream, the wave would experience the exact same degree of sound level decay with others during transmission and arrive maintaining designated arrival level differences. Thus, we construct F_i, a filter for each player, as a set of $L_i^D(f)$ for $\forall f$. At this moment, since an audio stream is represented in time domain, the filter of frequency domain should be transformed so as to be applied to the stream. So, the coordinator translates it into time domain by applying the inverse fast Fourier transform.

Continuous Audio Coordination

In conventional acoustic environments, it is assumed that users listen to music on a pre-defined sweet spot, played by loudspeakers in fixed locations. In mobile scenarios, however, such an assumption is unrealistic since both listeners and playing devices can be freely moved. Thereby, AMAC introduces a continuous audio coordination for better user experiences in dynamically changing sound environments. This continuous coordination is also useful to compensate for possible clock drifts caused by different clock speeds of devices.

Continuous acoustic measurement. As previously described, AMAC makes use of an acoustic control signal to capture acoustic characteristics before the playback starts. Although this measurement is essential for coordination, the signal currently used is undesirable for continuous measurement due to its noisy sound. As a solution to minimize such interferences, AMAC uses an inaudible acoustic signal for continuous coordination. First, the signal is modulated into an inaudible range of frequencies [19.2k, 24kHz]. A player merges the inaudible signal with its audio stream when the continuous coordination is required. At this time, to send out the signal more clearly, the player attenuates its stream by applying a low-pass filter with a cutoff frequency, 19.2kHz. A coordinator then extracts the inaudible signal from a transmitted sound by filtering out low frequencies [0, 19.2kHz], and computes impulse response of each player.

Continuous sound arrival time synchronization. Based on the acoustic response measurement, AMAC synchronizes sound arrival time of each player by delaying or skipping the audio stream. Let $T_i^D(k)$ denote a propagation delay of each player P_i for k-th measurement, where $T_i^D(0)$ is an initial propagation delay. At the initial coordination stage, AMAC compensates for different propagation delays of each player by computing the playback start time T_i^S based on $T_i^D(0)$. To keep the initially synchronized sound, additional compensation time T_i^C is determined as

$$T_i^C = T_i^D(k) - T_i^D(k-1).$$

Finally, P_i adjusts its audio stream based on T_i^C; it inserts zeroed data into the playback sample when T_i^C is negative (i.e., the stream is faster than the initial), or skips the sample by some amount when T_i^C is positive.

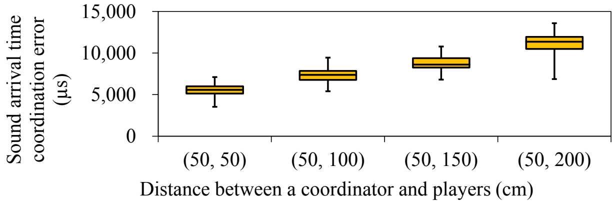
Continuous sound level equalization. While using inaudible signal does not affect the accuracy of sound arrival time synchronization, it does influence the correctness of sound level equalization. Since inaudible signals have some portions of frequencies filtered out, AMAC cannot accurately measure acoustic effects such as sound diffraction and reflection. Thus, unlike initial coordination which compensates various sound environments, AMAC is forced to only consider the distance attenuation for continuous coordination. At every coordination, a distance between a coordinator and each player is measured based on the sound arrival time [26]. Let $D_i(k)$ denote a distance of P_i measured at the kth continuous coordination, where $D_i(0)$ is an initial distance. According to the inverse-distance law, the compensation volume V_i^C is computed as $D_i(k)/D_i(0)$. The volume is then normalized to the maximum volume among players, because a simple amplification of sound level can cause a distortion. To this end, P_i applies the volume to its audio stream, and solves the distance attenuation during playback. Note that AMAC can achieve the more equalized sound if it uses audible signals. In the evaluation section, we describe such tradeoffs in more detail.

DEPLOYMENT

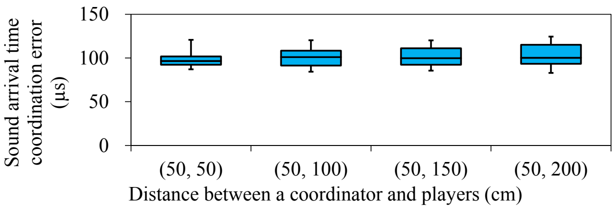
We implemented a prototype of AMAC on commodity smartphones (e.g., Nexus 4, Nexus S, and Nexus 10) running Android 4.2.2. In addition, we implemented a MMA application, MobileTheater, on top of AMAC API.

System Configuration. Each player transmits the MLS signal at a rate of 48kHz. As discussed, to accurately capture the sound transfer process, the MLS signal is 1.3 second long with 64k 16-bit signed short integers in a full range of frequencies, [0, 24kHz]. On the other hand, for continuous coordination, the inaudible MLS signal is used with the playback time of 26ms having a limited range of frequencies, [19.2, 24kHz]. A coordinator records the sound at a rate of 48kHz by using the MediaRecorder API of Android. From the measured signals, the coordinator constructs FIR filters that have 512 points for sound level equalization.

MobileTheater. MobileTheater playbacks a movie with 5.1 surround sound. It basically runs with six devices: five audio players and one video player. The video player takes a role of the coordinator while playing a video stream. MobileTheater synchronizes time and equalizes sound level through AMAC before starting playback. In the initialization stage, each player is manually assigned its own speaker role to play (e.g., center, left-front, or right-rear). The coordinator then splits an audio source into 5 different audio streams according to the roles of the players, and sends the audio streams to corresponding players. The players start actual playback at designated start times with continuous coordination. At this time, since players starts playback with some delay, 100ms, to avoid unpredictable latency, the video player also starts the



(a) NTP-based Baseline



(b) AMAC

Figure 4: Coordination accuracy in sound arrival time over different distances. Numbers in parentheses are the distance of fixed players and D^C respectively.

same amount later than its given start time. This way, MobileTheater provides 5.1 surround sound in coordination under AMAC.

EVALUATION

As explained, AMAC overcomes little support of commodity mobile devices and is successfully deployed as a prototype. In this section, we evaluate the capability of AMAC enabling MMA applications to reproduce high quality sound using the prototype and MobileTheater, with following metrics.

- *Coordination Accuracy.* How accurately does AMAC coordinate sound arrival times and levels? Do AMAC operate well in various environment, with dynamically-changing position of players?
- *User Experience.* How well do the sounds be coordinated on human's ears? Do MMA applications based on AMAC work in practice?

Experiment environments. We conducted experiments with one coordinator (Nexus 10, Samsung, 2012) and five players with two types (Nexus 4, LG, 2012 and Nexus S, Samsung, 2010). Since all the devices (Nexus 10, Nexus 4, Nexus S) are open source Android smartphones, we were able to run AMAC on them with some modifications of Android (e.g., *StableSoundPool* instead of *SoundPool*).

Most experiments were conducted in a classroom of 6.8m by 7.5m by 2.5m except for two cases. An experiment to see the effect of various room acoustics was taken in four different public areas. Also, user studies are performed in a park and the classroom, which are common places to listen to music with loudspeakers.

Experiments for Coordination Accuracy

In order to measure the coordination accuracy of AMAC, we used 5.1 channel audio streams with six devices. We forced a sound from a player P_i to be arrived at the coordinator, exactly 3 seconds after the sound from P_{i-1} is arrived, with the same sound arrival level. In this way, we measured the difference between sound arrival time and level of each player, and calculated their coordination errors whether they have designated time and level differences (i.e., 3 seconds and 0dB).

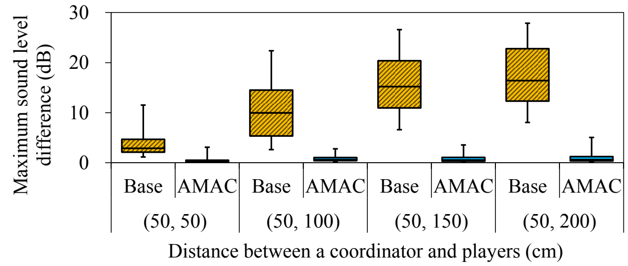


Figure 5: Coordination accuracy in sound arrival level over different distances.

For sound arrival time coordination, the comparable mechanism used is NTP, the most popular network-based time synchronization protocol. For each player P_i, we measured the sound arrival time error over designated time, and took the maximum error among all players as a result. We then created box-and-whisker plot with a dataset formed over 50 trials, using the 5th and 95th percentile as the ends of whiskers.

For sound arrival level coordination, since there exists no proper comparable mobile system, we considered commodity phones (AMAC without sound level coordination) as the baseline. Sound level measurements were performed in the range of frequencies [1k, 15kHz], since there are unplayable ranges [0, 1kHz] of commodity mobile devices within the range of human audible frequencies [0, 15kHz]. For each player P_i, we measured its sound arrival level, and gained the maximum error by subtracting the smallest level from the largest level among players. We then compiled statistics of errors over all frequencies [1k, 15kHz] to obtain a graph.

We show that AMAC provides a highly accurate coordination on sound arrival time and level, adapting to various, dynamically-changing mobile environments.

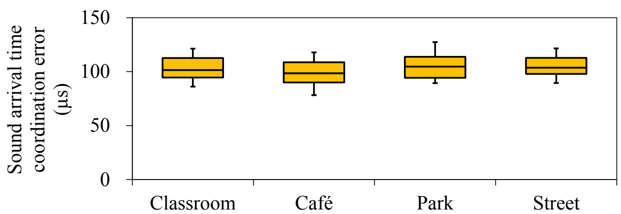
Irregular device layout. This experiment shows the effect of irregular layouts to the sound arrival time and level coordination accuracy, by controlling distances between each player and the controller. Among five players, two players were fixed 50cm away from the coordinator, and the other three players were placed at a certain distance D^C away from the coordinator. We conducted experiments with four layouts of different D^C's (50cm, 100cm, 150cm, and 200cm), comparing NTP to AMAC for time coordination and the baseline to AMAC for level coordination.

Figure 4 indicates that the sound arrival time coordination error of NTP is steadily increasing as D^C increases, in contrast to the stable accuracy of AMAC. Since the acoustic signal notifies the propagation delays to AMAC, it can adaptively adjusts each player's playback start time. It is worth noting that in all cases, AMAC meets the timing requirement (i.e., 1ms) for human auditory perception.

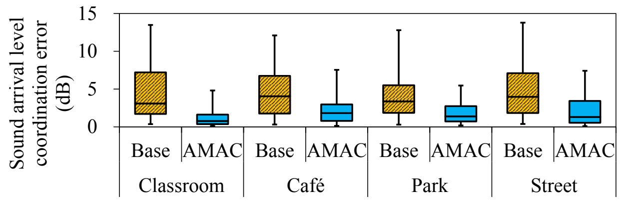
Figure 5 represents the coordination accuracy in sound arrival level over irregular layouts. Since we used devices of the same type (Nexus 4) which has similar acoustic characteristics, unintended effects from heterogeneous loudspeakers

Indoor		Outdoor	
Place	Noise (dBA)	Place	Noise (dBA)
Classroom	[30.5, 48.8]	Park	[44, 67]
Cafe	[64.9, 73.3]	Street	[67.1, 80.4]

Table 1: Four places for real-world experiments: two indoor and two outdoor places with different noise levels. The indoor places have many reflections while the outdoor places are open space which is free from the reflections.



(a) Accuracy of sound arrival time coordination



(b) Accuracy of sound arrival level coordination

Figure 6: Audio coordination in real environments

can be removed. As the figure shows, the error of baseline keeps growing as D^C increases while AMAC presents quite constant errors (e.g., decreased by 8.4-22.8dB at the 95th percentile). These results show that AMAC supports more accurately localized sound for human, even in irregular layouts.

Different room acoustics. This experiment evaluates how robust to room acoustics AMAC is in various real-world environments. We conducted experiments on four different public places as shown in Table 1. In order to observe the effect of various noise levels and reflections, we chose a quiet place and a noisy place for both indoor and outdoor environments. Since some places did not have enough space, we created a small layout; two players were located 30cm away from the coordinator, while the other three had different distances (50cm) to the coordinator.

Figure 6(a) shows that AMAC consistently provides the sound arrival time accuracy of less than 120μs of error across all the places. We note that NTP-based coordination never meets a desired accuracy of 1ms of error in the experiments (the results are not shown due to page limitation). Figure 6(b) represents much better sound level coordination of AMAC compared to the baseline. The figure shows that the coordination error of AMAC in noisy places slightly increases compared to quiet places (2.73dB for indoor and 1.96dB for outdoor at the 95th percentile). This indicates that AMAC is quite robust to noise in many common public places.

Device heterogeneity. In this experiment, we show the effectiveness of AMAC in sound arrival time and level coordination on different devices. We used three Nexus 4's and two Nexus S's for players, and placed them at the same distance (50cm) from the coordinator to solely focus on the device heterogeneity issue. Note that as we figured out device heterogeneity did not affect the sound arrival time accuracy much, we do not include the case due to space limit.

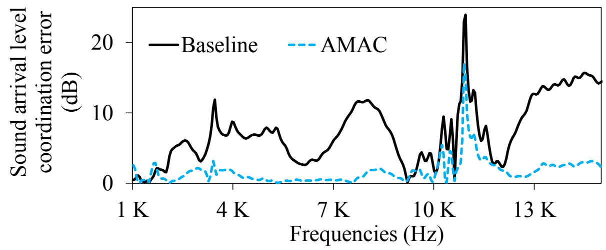


Figure 7: Coordination accuracy in sound arrival level over device heterogeneity.

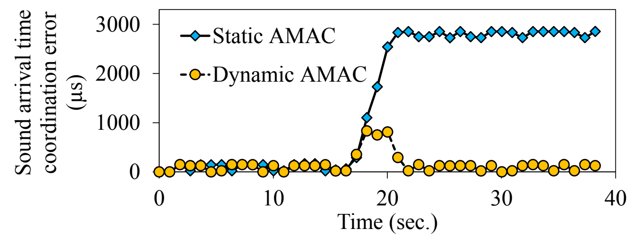


Figure 8: Coordination accuracy of sound arrival time coordination in dynamic loudspeaker layouts.

Figure 7 shows how the sound level error varies from 1kHz to 15kHz, comparing the baseline (i.e., without sound level coordination) to AMAC. In the figure, the baseline shows a large degree of error in most frequencies with the median of 6.79dB, contrary to much lower errors, the median of 1.67dB, of AMAC. Notice that there exists a large error shown at the frequency of 10.9kHz in both the baseline and AMAC. As discussed in the sound arrival level coordination section, this large error is caused by the limitation of loudspeakers embedded in mobile devices, and such deep valleys are disregarded in AMAC to avoid an over-attenuation. Thus, AMAC provides far greater accuracy in coordinating sound arrival level over device heterogeneity but for the exceptional valley.

Dynamic sound environments. So far, we conducted experiments on static environments. This experiment is to observe how effectively AMAC deals with dynamically-changing situation, such as device location change during playback. In the experiment, we initially located five homogeneous players (Nexus S) with the same distance (50cm) to the coordinator. A few seconds after the audio stream played, we slowly pushed one of the players away (from 17s to 20s and from 50cm to 150cm) to simulate user mobility.

For continuous time coordination, we measured changes of the coordination errors during total playback, comparing dynamic AMAC (i.e., coordinate every 1 second) to static AMAC (i.e., coordinate once before playback). Figure 8 shows that the sound arrival time error of static AMAC increases as the player moves, and it does not drop after the movement. The final error of 3ms is far larger than the timing

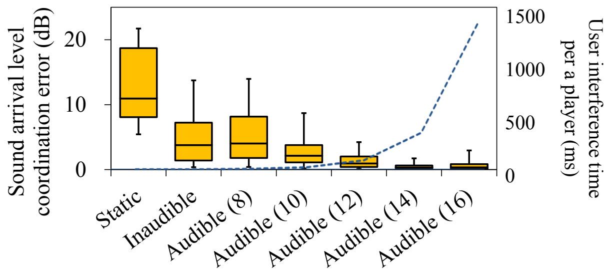
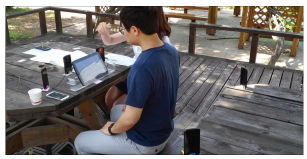


Figure 9: Coordination accuracy of sound arrival level coordination in dynamic loudspeaker layouts. Parentheses means the order of MLS signal.

(a) Indoor (classroom).



(b) Outdoor (park)

Figure 10: The experiment places of our user study.

requirement of 1ms. In contrast, dynamic AMAC represents nearly zero coordination error after the player stops moving, while it also meets the timing requirement even during the movement.

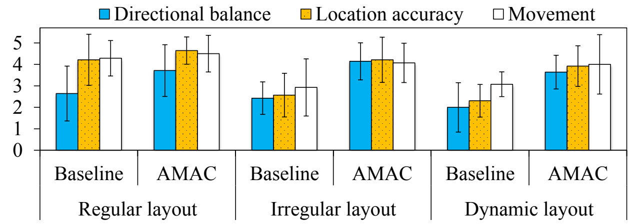
Figure 9 shows the effect and user interference of continuous level coordination over different acoustic signals, compared to static AMAC. We measured sound arrival level coordination errors after the movement of a player stops. Since static AMAC keeps using a filter constructed at the initial time, it suffers from high coordination error. For dynamic AMAC, the figure shows that the level coordination error decreases as the order of acoustic signal increases (i.e., emits longer signal), while user interference time exponentially increases. Dynamic AMAC with an inaudible signal also shows quite better performance than static AMAC while not disturbing users. However, it still has some non-negligible coordination errors, compared to AMAC with audible signals. Thus, AMAC requires further optimization to dynamically reduce the level coordination errors, without user interference.

User Study

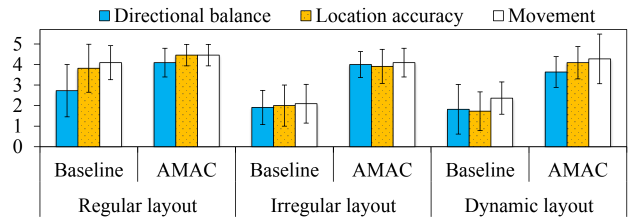
We conducted a user study to evaluate the effect of AMAC on users while experiencing MMA applications. The objective of this study was to observe how AMAC affect the users' sound perception in various acoustic environments.

User study design. Thirty university participants (11 females) aged 21 to 29 years were selected to participate in this experiment out of sixty applicants, since the participants showed greater interests in music listening and video watching. Each participant was asked to use a MMA application, MobileTheater, to listen to an instrumental song and watch a video clip with 5.1 surround sound. MobileTheater ran on five players (Nexus 4) and one coordinator (Nexus 10).

To encompass various acoustic environments, we conducted the experiment with three different loudspeaker layouts and two different places. We chose three representative layouts: (i) regular where all the players were located with the same distance (50cm) from the user, (ii) irregular where two players were placed with a further distance (100cm), and (iii) dynamic where two players were moved during the playback. As shown in Figure 10, we selected two common places: (i) a classroom as an indoor place with a moderate degree of sound reflection but no public noise, and (ii) a park as an outdoor, open place with little reflection but with more noise in public ([48.6, 75.9dBA]). Each participant experienced the experiment for all the three layouts in either place (15 in the classroom and the other 15 in the park).



(a) Indoor (classroom).



(b) Outdoor (park)

Figure 11: User experiences: spatial impression of sound.

Since there is no comparable system for sound arrival level equalization, we compared AMAC to the baseline with NTP clock synchronization and no level equalization. Participants were designed to experience two different types of sound listening mode, with and without visual images. In the instrumental music case, all devices play the same audio stream with the same speaker configuration, and participants were asked to focus on directional balance of the sound. In the video clip case, the video showed objects moving around and presented 3D sound effects accordingly by adjusting the sound level of each speaker, concentrating on locational accuracy and movement. After each experiment, participants were requested to fill in a 5-point scale questionnaire, covering directional balance, locational accuracy, and movement. It was written based on ITU-R recommendation [16], which suggests well-known evaluation methods of sound system.

User study results. Through this experiment, we were able to confirm some positive benefits of AMAC in reproducing the sound with spatial impressions. Figure 11(a) compares the users' spatial impressions with AMAC to the baseline, across three different device layouts and two different places. In the regular layout case, the baseline is able to deliver audio streams in some comparable sound levels enough to create similar location accuracy and movement impressions. However, it is not able to maintain sound arrival times accurately, and since an auditory event is dominantly localized by the first-arriving sound, this leads to creating some biased directional balance ($p < 0.03$). In the irregular layout case, the baseline is no longer able to maintain designated sound arrival levels due to the different distances of devices. This results in poor performance in producing accurate location and movement impressions, compared to AMAC ($p < 0.001$). In the dynamic layout case, AMAC is shown to maintain the spatial impressions through continuous coordination, while the baseline shows relatively poor performance.

Even though some participants had an uncomfortable feeling in initialization process of AMAC (i.e., the noisy MLS signal sound and about 30 seconds of initialization for six devices), most of users were impressed by the benefits of AMAC with the following comments: "a brand new experience that I've never experienced before", "feel like using better speakers", and "the sound was exactly consistent with an object moving". One of the participants in the outdoor experiment said "I didn't even notice the (public) noise until mentioned".

RELATED WORK

Audio engineering. The past few decades have seen a proliferation of techniques in acoustic engineering to improve sound reproduction quality with multiple loudspeakers (see [2, 33] for more details). For example, sound spatialization has been introduced to control the perceived direction of an auditory event for stereo sound systems, basically adjusting the sound level difference between two loudspeakers (i.e., amplitude panning) based on the stereophonic law of sines [3] and tangents [4]. The amplitude panning techniques are extended for two or more-channel surround sound systems [29]. Those amplitude panning methods commonly require that homogeneous loudspeakers be placed equally distant from listeners in a sweet spot for optimum sound performance. As such, under the implicit assumption of homogeneous and stationary speakers, most traditional acoustic engineering techniques paid little attention to adapting to different, changing situations, in which speakers are heterogeneous and placed arbitrarily under various room acoustics conditions.

Nowadays many audio systems [1, 27, 31] provide off-line calibration functions of loudspeaker layouts, and the development of the state-of-the art equalization technique is not the ultimate goal of this study. Although room response equalization techniques are quite well-known [25, 19], most of the equalization techniques are not suitable for the mobile listening scenario, in which the individual devices with nonidentical clock drifts are connected wirelessly and both the listener and mobile devices can move during the audio playback. Previous example dealing with continuous time-synchronization and equalization can be found from the work of Härmä [15], in which individual gains and time delays between two loudspeakers are continuously identified by comparing subband-envelopes of the source and measured sound signals. Nevertheless, the matching technique is in question for the multi-channel reproduction scenario for which multiple loudspeakers are activated at the same time and produce a nonidentical subband envelope at the listener position.

Clock synchronization. Traditionally, many researchers have shown interests to synchronize devices in the wireless network (see [32] for a survey). NTP [23] is the most popular network-based time synchronization protocol. By exchanging packets between a sender and receivers, the clock of receivers is adjusted to the received time. This scheme works if the delay between sending and receiving packets is negligible. However, due to the unpredictable nature of wireless communication, NTP suffers from low precision of tens of milliseconds. Others [12, 13] propose more precise (a few μs) synchronization techniques. RBS [12] discards the non-deterministic of the send time by using a receiver-receiver synchronization algorithm. On the other hand, TPSN [13] maintains the sender-receiver algorithm, but overcomes send time errors using time stamps at MAC layer. If all devices have the same propagation delay, such network-based synchronization can achieve highly accurate coordination for the sound arrival time. However, in realistic cases, irregular loudspeaker layouts and different room acoustics lead a large difference of the propagation delays. To this end, since the network-based synchronization does not give any compensation for such difference, it fails to satisfy the desired timing requirement.

Acoustic localization. Acoustic signal has been widely used in mobile computing for distance measurement. BeepBeep [26] introduces the first use of acoustic signals for phone-to-phone distance measurement. SwordFight [34] improves accuracy and frequency for the support of mobile motion games. The signal also helps to provide centimeter resolution on infrastructure-based localization [22]. While such studies focus on propagation delay for localization, this work also considers sound level decay to sense the acoustic characteristics of mobile listening environments.

CONCLUSION

In this paper, we describe the design and implementation for AMAC, which supports a new sound system abstraction that hides acoustics-specific details in mobile listening environments. The key enabling techniques for this new abstraction is a highly-accurate adaptive sound coordination in both time and frequency domains, which provides APIs on top of which a new emerging class of MMA applications, such as MobileTheater, can be built. Our experience with real users is positive. MobileTheater on AMAC delivers a rich, immersive listening experience. However, our work is a starting point for a number of research directions. For example, AMAC employs inaudible acoustic signals for continuous coordination while incurring no user interference. However, the use of inaudible signals introduces some non-negligible coordination errors, compared to the case where long audible signals are used yet at the expense of disturbing users. Our future work includes investigation of techniques to reduce coordination errors without degrading user experience.

ACKNOWLEDGMENTS

This work was supported in part by KEIT (2011-10041313), KIAT (M002300089), and NRF (2010-0020729) funded by the Korea Goverment (MSIP/MOTIE).

REFERENCES

1. Audyssey Laboratories, Inc. MultEQ. http://www.audyssey.com/technologies/multeq.

2. Baalman, M. A. J. Spatial Composition Techniques and Sound Spatialisation Technologies. *Organised Sound 15*, 3 (2010), 209–218.
3. Bauer, B. B. Phasor analysis of some stereophonic phenomena. *The Journal of the Acoustical Society of America 33*, 11 (1961), 1536–1539.
4. Bennett, J. C., Barker, K., and Edeko, F. O. A new approach to the assessment of stereophonic sound system performance. *Journal of the Audio Engineering Society 33*, 5 (1985), 314–321.
5. Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, revised edition, 1997.
6. Borish, J., and Angell, J. B. An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise. *Journal of the Audio Engineering Society 31*, 7/8 (1983), 478–488.
7. Breebaart, J., Engdegard, J., Falch, C., Hellmuth, O., Hilpert, J., Hoelzer, A., Koppens, J., Oomen, W., Resch, B., Schuijers, E., and Terentiev, L. Spatial Audio Object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding. In *Proc. of the 124th Proceedings of the 124th convention of the Audio Eng. Soc.* (2008).
8. Cheer, J., Elliott, S. J., Kim, Y., and Choi, J. W. Practical implementation of personal audio in a mobile device. *Journal of the Audio Engineering Society 61*, 5 (2013), 290–300.
9. Choi, J. W., and Kim, Y. H. Generation of an acoustically bright zone with an illuminated region using multiple sources. *Journal of the Acoustical Society of America 111*, 4 (2002), 1695–1700.
10. Choi, J. W., and Kim, Y. H. Integral approach for reproduction of virtual sound source surrounded by loudspeaker array. *IEEE Transactions on Audio, Speech and Language Processing 20*, 7 (2012), 1976–1989.
11. Cohn, M., and Lempel, A. On Fast M-Sequence Transforms. *IEEE Transactions on Information Theory 23*, 1 (1977), 135–137.
12. Elson, J., Girod, L., and Estrin, D. Fine-grained network time synchronization using reference broadcasts. *ACM SIGOPS Operating Systems Review 36*, SI (2002), 147–163.
13. Ganeriwal, S., Kumar, R., and Srivastava, M. B. Timing-sync protocal for sensor networks. In *Proceeding of the 1st international conference on Embedded networked sensor systems* (2003).
14. George, N. V., and Panda, G. Advances in active noise control: A survey, with emphasis on recent nonlinear techniques. *Signal processing 93*, 2 (2013), 363=377.
15. Harma, A. Online acoustic measurements in a networked audio systems. In *Proceedings of 120th convention of Audio Enginnering Society* (2006).
16. International Telecommunication Union. *ITU-R BS.1284-1: General methods for the subjective assessment of sound quality*.
17. International Telecommunication Union. *ITU-R BS.775-3: Multichannel stereophonic sound system with and without accompanying picture*.
18. Kim, Y. H., and Choi, J. W. *Sound Visualization and Manipulation*. Jonh Wiley & Sons, 2013.
19. Kuriyama, J., and Furukawa, Y. Adaptive loudspeaker system. *Journal of the Audio Engineering Society 37*, 11 (1989), 919–926.
20. Lee, Y., Min, C., Hwang, C., Lee, J., Hwang, I., Ju, Y., Yoo, C., Moon, M., Lee, U., and Song, J. Sociophone: Evereday face-to-face interaction monitoring platform using multi-phone sensor fusion. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services* (2013).
21. Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. The precedence effect. *The Journal of the Acoustical Society of America 106*, 4 (1999), 1633–1654.
22. Liu, K., Liu, X., and Li, X. Guoguo: Enabling Fine-grained Indoor Localization via Smartphone. In *Proceedings of the 11th annual international conference on Mobile systems, applications, and services* (2013).
23. Mills, D. L. Internet Time Synchronization: The Network Time Protocol. *IEEE Transactions on Communications 39*, 10 (1991), 1482–1493.
24. Mills, D. L. Network Time Protocol Version 4 Reference and Implementation Guide. Tech. rep., Electrical and Computer Engineering, 2006.
25. Miyoshi, M., and Kaneda, Y. Inverse filtering of room acoustics. *IEEE Transactions on Acoustics, Speech and Signal Processing 36*, 2 (1988), 145–152.
26. Peng, C., Shen, G., Zhang, Y., Li, Y., and Tan, K. BeepBeep: A High Accuracy Acoustic Ranging System using COTS Mobile Devices. In *Proceedings of the 5th international conference on Embedded networked sensor systems* (2007).
27. Pioneer. MCACC. http://www.pioneerelectronics.com/pio/staticfiles/index.html.
28. Pomberger, H., and Zotter, F. An Ambisonics format for flexible playback layouts. In *Proceedings of the 1st Ambisonics Symposium* (2009).
29. Pulkki, V. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society 45*, 6 (1997), 456–466.
30. Samsung. GroupPlay. http://content.samsung.com/us/contents/aboutn/groupPlay.do.
31. Sony. Digital Cinema Auto Calibration (D.C.A.C). http://www.sony-mea.com/article/272779/section/product/product/bdv-it1000.

32. Sundararaman, B., Buy, U., and Kshemkalyani, A. D. Clock synchronization for wireless sensor networks: a servey. *Ad Hoc Networks 3*, 3 (2005), 281–323.

33. Toole, F. *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms*. Taylor & Francis, 2009.

34. Zhang, Z., Chu, D., Chen, X., and Moscibroda, T. SwordFight: Enabling a New Class of Phone-to-Phone Action Games on Commodity Phones. In *Proceedings of the 10th international conference on Mobile systems, applications, and services* (2012).