IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 1, 2019, accepted October 17, 2019, date of publication October 31, 2019, date of current version November 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950780

Audio Effect Units in Mobile Devices for Electric Musical Instruments

SHU-NUNG YAO, (Member, IEEE)
Department of Electrical Engineering, National Taipei University, New Taipei City 23741, Taiwan
e-mail: snyao@gm.ntpu.edu.tw

This work was supported in part by the Ministry of Science and Technology in Taiwan under Grant MOST 107-2221-E-305-010-MY2.

ABSTRACT With the advent of modern techniques, there has been an increase in mobile devices with powerful functions. This study aims to utilize application software instead of hardware equipment to alter the sound of electric musical instruments by developing the functions of audio effect units on a mobile device. The built-in software audio effects include dynamic effects, delay effects, mixing, and equalization. Multichannel audio signal processing is proposed to enhance sound externalization. The listening tests indicate that the sound effects together with a quadraphonic system produce superior special effects and spatial audio than conventional effect units. Additionally, a denoiser is proposed for noise reduction, which is a function especially helpful for singers performing on busy streets. The customization in the module is achieved by a personal voice preprocessing system. A dual-filter function was utilized for the denoiser to adjust to different environments. The objective performance measurements demonstrate that the proposed denoiser outperforms the state-of-the-art methods. A hardware audio interface serving as the connection between a mobile device and an electric instrument, such as an electric guitar, electric piano, or electric bass, was also built, providing impedance matching and voltage balance. The proposed interface transfers the electric instrument signal into a 3.5-mm jack in a smartphone or tablet. The developed audio interface is light and low-noise and can operate without being connected to an external power supply, thereby making it suitable for street musicians. The experiments validate the feasibility of using the proposed circuit for real-time audio signal conversion.

INDEX TERMS Audio effects, electric bass, electric guitar, multichannel audio, noise reduction.

I. INTRODUCTION

On a busy street, there are often buskers performing. Street singers are street performance artists who perform by singing and playing musical instruments. If they play and sing on their own, a piano or a guitar is usually the favorite instrument. Compared with a piano, a guitar has better mobility, which is essential in street performance. However, even if the performers accompany themselves on the guitar with effect units, direct interface units, amplifiers, and loudspeakers, much equipment that street musicians use is inconvenient, especially in changeable weather. In this work, immigrating software sound effects into a tablet, smartphone, or notebook, thereby reducing the weight of professional audio equipment, is proposed. In addition, an audio interface providing a connection between an electric instrument and a mobile device was developed. The proposed system showed better audio quality than commercial products.

The associate editor coordinating the review of this manuscript and approving it for publication was Huawei Chen.

In the early stages, sound effect units were normally implemented by analog circuits. With the advent of digital techniques, a digital signal processor (DSP) or a field-programmable gate array (FPGA) provided sound engineers with another choice. Ling *et al.* [1] implemented special sound effects on a DSP, including reverberation by a first-order infinite impulse response (IIR) filter and pitch control by frequency shifting. Byun *et al.* [2] designed sound effects on both a DSP and an FPGA, achieving a system on a chip. Whereas most sound effects were implemented on a DSP chip, equalization was implemented on FPGA hardware since the equalizer required a large amount of computational power. Liu *et al.* [3] proposed a low-power FPGA-based structure for delay effects. Mohamadi [4] designed an audio system on a sound card paired with a computer. Although several reports on designs of the system using digital circuits have been published, there have been

FIGURE 1. Connecting the wires of the proposed system.

few studies of software-based implementations. Whereas Anghelescu *et al*. [5] used the C# language to develop special sound systems on a desktop, an audio system for mobile devices was designed in this study. Moreover, unlike Anghelescu *et al*. [5], who applied an existing software library, multichannel sound effects and a denoiser with data preprocessing algorithms, which are rarely found in existing sound effects, were invented in the present work. An example of the system layout is illustrated in Fig. 1. An electric guitar is connected to the developed audio interface for transferring the signal to a mobile device. The proposed software audio units on the mobile device alter the sound with special effects and subsequently transmit it to a loudspeaker via the audio interface. The proposed audio units can also transmit the sound to a loudspeaker array by utilizing a multichannel sound card, thereby creating an all-enveloping sound experience.

The remainder of this paper is organized as follows. Section II introduces the proposed sound effects and the conventional effect units. A user-friendly graphical user interface (GUI) is used for improving the dynamic effects. Multichannel surround sound is employed for the delay effects to enhance auditory spatial perception. The proposed mixer provides fine frequency resolution and a wide range of listening environments. The denoiser is proposed to improve the vocal quality. Section III presents the circuit design of the proposed audio amplifier that converts a high-impedance audio signal to a low-impedance one. Objective and subjective tests for the proposed sound effects and audio amplifier are provided in Section IV, and the conclusions and directions for future work are presented in Section V.

II. EFFECT UNITS

Street singers and audio engineers use effect units during live performances and in the studio, respectively. Although effect units typically accompany the electric guitar, electronic keyboard, electric piano, and electric bass, effects can also be used with acoustic instruments, such as drums and vocals, in unplugged live performances.

Analog electronics are still used for special sound effects nowadays, but digital equivalents have become far more common. Even if the effect units are digitalized, street musicians have to bring many effect boxes, such as stompboxes, rackmounts, and tabletop units, to a live show. This is especially inconvenient for a single street musician. Therefore, the use of software instead of hardware is proposed. A few of the most popular types of effect unit were expended and modified in this work so that the treated sound can be more special and spatial than that of a conventional effect unit. A denoiser, which is not standard equipment in an effect rack, was also built. All the proposed effect units were designed using software plug-ins for laptops and mobile applications for mobile devices such as smartphones and tablets.

A. DYNAMIC EFFECTS

The dynamic effect only affects the gain or the volume of the sound. The most common dynamic effects are compressors and limiters. A compressor does not change the loudness of the sound when the input volume is less than a user-defined threshold. If the input volume exceeds the threshold, the input signal is attenuated to a lesser degree, compressing the dynamic range of a signal, as shown in Fig. 2a. When the input volume exceeds the threshold, the limiter makes the output level equal to the threshold, as shown in Fig. 2a. When dynamic effects are used on vocal tracks, the voice maintains a more consistent level. For guitar tracks, dynamic effects prevent an excessively loud sound when the guitar is violently strummed. In traditional stompboxes, guitar players adjust the threshold and dynamic range by means of push knobs, a design which is not user-friendly nor intuitive for novices. In the GUI presented, the relationship

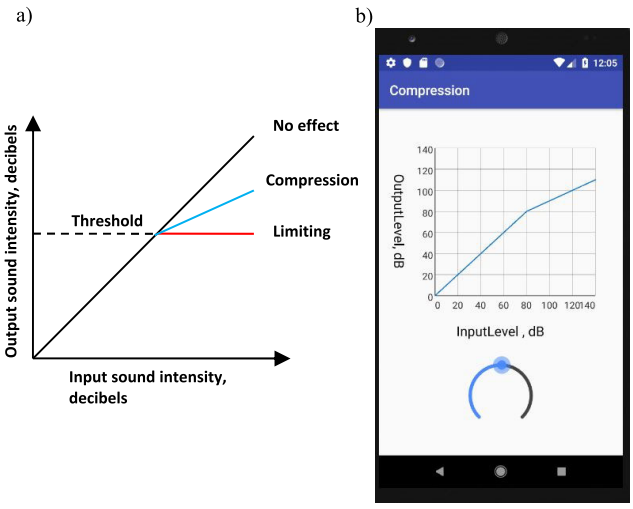

FIGURE 2. Dynamic effects including: (a) compressor and limiter. (b) Screenshot of dynamic effects in the proposed system.

between the output level and the input level is plotted on a screen, as shown in Fig. 2b, to make the adjustment more understandable. In the proposed system design, the threshold can be automatically adjusted. The user recorded a vocal piece for the system to calculate the decibel level. To avoid a masking effect, the threshold of a musical instrument is 3 dB lower than the voice level.

B. DELAY EFFECTS

A crude echo effect is composed of a time-delayed and attenuated input signal which is added to the original one. The simple echo effect does not generate a convincing sense that the audio was actually recorded in a real room with real echoes. Since there is normally more than one reflective surface in a real room, multiple echoes with different time delays can be perceived.

A more elaborate reverberation effect was therefore built. One way to implement realistic reverberation is to measure the impulse response of the space in the actual recording environment. However, it is inconvenient for most people to prepare the equipment and measure the impulse response of all the rooms or buildings they want to simulate. A more economical method of creating a reverberation effect can be through the use of an IIR filter [1] producing multiple delayed and fading replicas of the input signal.

In the proposed design, a room model is used to achieve a more realistic reverberation and can also be applied to a multichannel surround-sound system. The model is based on the image-source method [6], [7]. By adjusting the parameters, the effect unit can be easily reconfigured to simulate rooms of different sizes.

To render a full two-dimensional acoustic space for multiple listening positions, more than two loudspeakers are required. Quadraphonic audio was probably the earliest consumer product in surround sound, and the loudspeaker layout is simpler than the modern 5.1 or 7.1 home theater. Because of its convenience for street singers, a four-loudspeaker array was chosen to construct the room model and to improve the multichannel customer experience. There were several audio formats, such as discrete formats and matrix formats, developed for quadraphonic sound in 1970s [8]. In the performance scene, it is rarely easy to place loudspeakers in a rigid arrangement. Hence, Ambisonics, proposed by Gerzon [9], can be a suitable audio format for street singers owing to its flexibility of loudspeaker configuration.

A sound source U can be positioned anywhere in the two-dimensional space, as shown in Fig. 3. The Ambisonic signals associated with the location of U are generated by the following encoding equations:

$$\mathrm{W} = U \cdot 0.7071, \tag{1}$$

$$\mathrm{X} = U \cdot \cos\theta, \tag{2}$$

and

$$\mathrm{Y} = U \cdot \sin\theta. \tag{3}$$

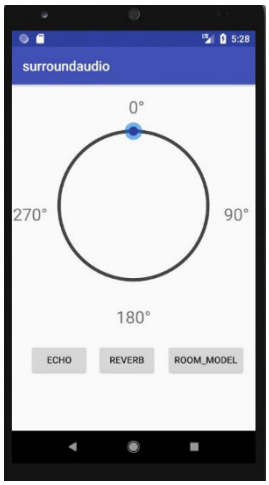

FIGURE 3. Screenshot of delay effects including echo, reverberation, and the proposed room model.

According to Ambisonics, the loudspeaker feed b_k is expressed as

$$b_k = \frac{1}{4}\left[W\left(\frac{1}{\sqrt{2}}\right) + X\left(\cos\vartheta_k\right) + Y\left(\sin\vartheta_k\right)\right], k = 1, 2, 3, 4 \tag{4}$$

where ϑ_k denotes the position of the k^{th} loudspeaker. If the loudspeaker array is in a square, $\vartheta_1 = 45°$, $\vartheta_2 = 135°$, $\vartheta_3 = 225°$, and $\vartheta_4 = 315°$. The design flow is shown in Fig. 4. The sound source U might be the vocal or signal from an instrument, and the room model calculates the delay and decay ($\mathrm{Delay_i}$ and $\mathrm{G_i}$) of the real source and the mirror source. The angle of the real source and the mirror source were encoded in Ambisonic signals by (1), (2), and (3). They are summed up to obtain the encoded components. Finally, (4) is used to render the sound field by a quadraphonic loudspeaker array.

When a mobile device is used, only a stereo output is available. If one wants to build a quadraphonic array, C-format (UHJ format) can be used. First, the W, X, and Y components are encoded into S and D components, as shown in

$$\mathrm{S} = 0.9396926\mathrm{W} + 0.1855740\mathrm{X}, \tag{5}$$

and

$$\mathrm{D} = \mathrm{j}\left(-0.3420201\mathrm{W} + 0.5098604\mathrm{X}\right) + 0.6554516\mathrm{Y} \tag{6}$$

where j is a $+\frac{\pi}{2}$ phase shift. Then, the left channel LC_s and the right channel RC_s are used in stereo format to transmit the sum and difference of the S and D channels as presented in

$$\mathrm{LC}_s = \frac{(\mathrm{S} + \mathrm{D})}{2}, \tag{7}$$

and

$$\mathrm{RC}_s = \frac{(\mathrm{S} - \mathrm{D})}{2}. \tag{8}$$

Originally, it is necessary to use three channels to transmit the W, X, and Y components, and those components can be

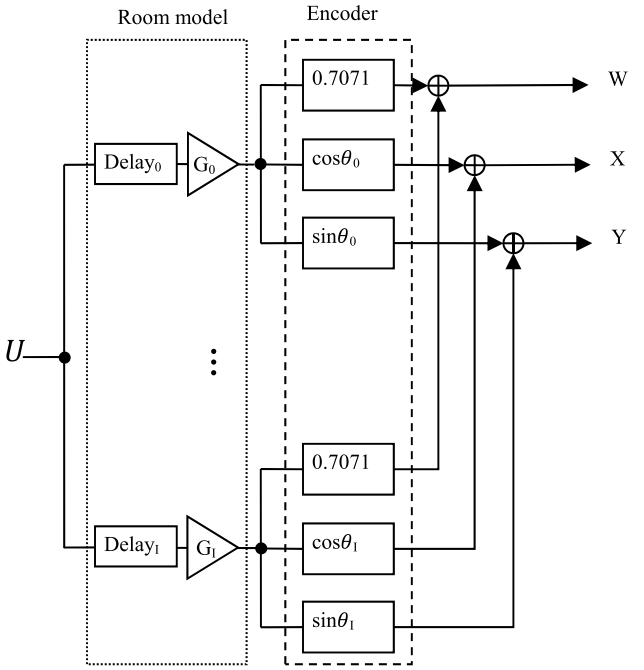

FIGURE 4. Room model with the first-order Ambisonic encoder. The number of mirror images is *I*. The angles inside the trigonometric functions depend on the positions of a real source and mirror sources.

encoded into two channels after using C-format. The first step of the decoding process is to extract the S and D components from the stereo audio, as shown in

$$\mathrm{S} = \mathrm{LC}_s + \mathrm{RC}_s, \tag{9}$$

and

$$\mathrm{D} = \mathrm{LC}_s - \mathrm{RC}_s, \tag{10}$$

and the reconstructed W, X, and Y components are

$$\mathrm{W}' = 0.982\mathrm{S} + \mathrm{j}(0.164\mathrm{D}), \tag{11}$$
$$\mathrm{X}' = 0.419\mathrm{S} - \mathrm{j}(0.828\mathrm{D}), \tag{12}$$

and

$$\mathrm{Y}' = 0.763\mathrm{D} + \mathrm{j}(0.385\mathrm{S}), \tag{13}$$

where W', X', and Y' are used to indicate that the components cannot be perfectly reconstructed. The $+\frac{\pi}{2}$ phase shifter was developed according to the previous work [10] by using the discrete-time Hilbert transform design. The mobile device and the four loudspeakers were connected by using a multichannel sound card, as shown in Fig. 5.

C. MIXER

The main purpose of a mixer is to add multiple signals together. However, before signal mixture, the sound in each audio channel can be independently modified. Fig. 6 shows the block diagram of a three-channel mixer. The signal path from each of the inputs enables the performer not only to vary the volume of each ingredient of the final stereo mix by a fader but also to modify its tonal color by equalization and its stereo image position by a pan control. A screenshot of the software mixer is shown in Fig. 7.

Equalization effects were developed to modify the frequency components of the input signals. An example found in many high-end audio systems is a loudspeaker crossover. It is a three-band equalization section, splitting an audio feed into high-, mid-, and low-frequency components separately routed

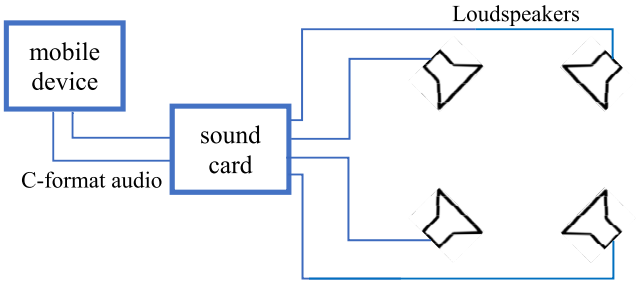

FIGURE 5. Using a mobile device to produce multichannel surround sound.

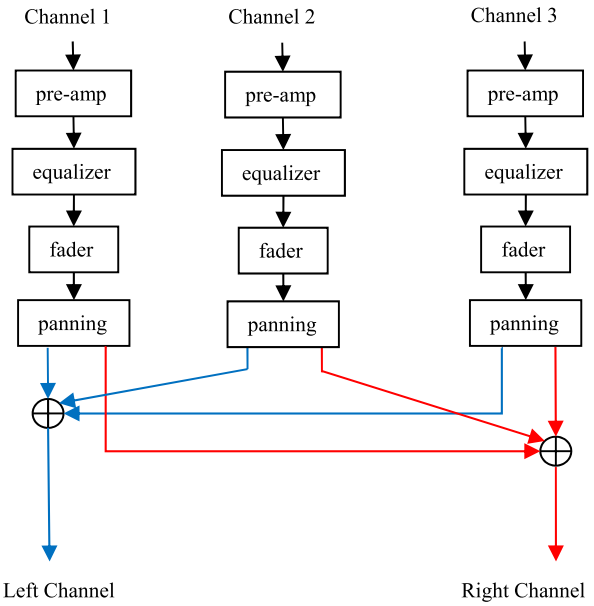

FIGURE 6. Block diagram of three-channel mixer.

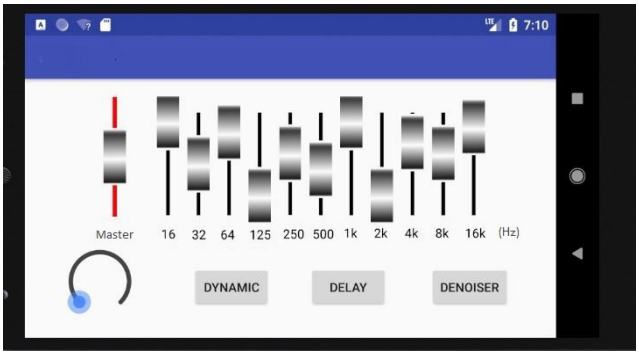

FIGURE 7. Screenshot of the proposed software mixer.

to tweeters, midranges, and woofers [11]. The equalizer can also compensate for a distorted signal [12]. In the proposed system, more-precise control of equalization parameters was designed. Using 11 frequency bands makes a much-finer control of the overall frequency response possible. Each frequency band can be amplified or attenuated as desired.

Unlike the conventional filterbank composed solely of band-pass filters, the proposed equalizer contains a low-pass filter, a high-pass filter, and nine band-pass filters. The digital filters in this study were derived from a hybrid method [13] which combines the use of a bilinear transformation and the pole-zero placement techniques and employs a maximum point normalization (MPN) method, to develop a simple design procedure. The coefficients of the filter transfer functions designed by the proposed method are identical to those obtained by the bilinear transformation. However, the proposed algorithm based on designing the 3-dB frequencies is simpler than the conventional method, based on designing the stopband and passband frequencies, in the previous research [14]. Unlike the pole-zero placement technique in previous works [15], [16], the proposed procedure is compatible with higher-order digital filter design.

Suppose that an L^{th}-order digital filter is in the form

$$H\,(\mathrm{z}) = G\frac{B(\mathrm{z})}{A(\mathrm{z})} = GH^{'}\,(\mathrm{z})\,, \tag{14}$$

where G is a gain factor, $B(\mathrm{z})$ is the numerator, and $A(\mathrm{z})$ is the denominator. If the denominator is for a low-pass filter or a high-pass filter, it can be expressed as

$$A\,(\mathrm{z}) = \left(1 + \mathrm{a}_{01}\mathrm{z}^{-1}\right)\left(1 + \mathrm{a}_{11}\mathrm{z}^{-1} + \mathrm{a}_{12}\mathrm{z}^{-2}\right) \ldots \left(1 + \mathrm{a}_{\mathrm{J}1}\mathrm{z}^{-1} + \mathrm{a}_{\mathrm{J}2}\mathrm{z}^{-2}\right), \tag{15}$$

where J is the largest integer not greater than $L/2$ and

$$\mathrm{a}_{01} = \frac{1-(-1)^{L}}{2}\frac{\Omega_0 - 1}{\Omega_0 + 1}, \tag{16}$$

$$\mathrm{a}_{\mathrm{i}1} = \frac{2(\Omega_0^2 - 1)}{1 - 2\Omega_0\cos\Theta_{\mathrm{i}} + \Omega_0^2}, \tag{17}$$

and

$$\mathrm{a}_{\mathrm{i}2} = \frac{1 + 2\Omega_0\cos\Theta_{\mathrm{i}} + \Omega_0^2}{1 - 2\Omega_0\cos\Theta_{\mathrm{i}} + \Omega_0^2}, \tag{18}$$

in which

$$\Theta_{\mathrm{i}} = \frac{(L+2\mathrm{i}-1)\pi}{2L}, \quad \mathrm{i} = 1, 2, \ldots \mathrm{J}. \tag{19}$$

Assuming that the 3-dB cutoff frequency is f_{c} and the sampling rate is f_{s}, then

$$\Omega_0 = \tan\left(\frac{\pi f_{\mathrm{c}}}{f_{\mathrm{s}}}\right). \tag{20}$$

The numerator B(z) is $(1+\mathrm{z}^{-1})^{L}$ for a low-pass filter and $(1-\mathrm{z}^{-1})^{L}$ for a high-pass filter.

In terms of digital band-pass filters, the general form of the denominator is

$$A\,(\mathrm{z}) = \left(1 + \mathrm{a}_{01}\mathrm{z}^{-1} + \mathrm{a}_{02}\mathrm{z}^{-2}\right)\left(1 + \mathrm{a}_{11}\mathrm{z}^{-1} + \mathrm{a}_{12}\mathrm{z}^{-2} + \mathrm{a}_{13}\mathrm{z}^{-3} + \mathrm{a}_{14}\mathrm{z}^{-4}\right) \ldots \left(1 + \mathrm{a}_{\mathrm{J}1}\mathrm{z}^{-1} + \mathrm{a}_{\mathrm{J}2}\mathrm{z}^{-2} + \mathrm{a}_{\mathrm{J}3}\mathrm{z}^{-3} + \mathrm{a}_{\mathrm{J}4}\mathrm{z}^{-4}\right), \tag{21}$$

where

$$\mathrm{a}_{01} = \frac{1-(-1)^{L}}{2}\frac{-2\mathrm{c}}{\Omega_0 + 1}, \tag{22}$$

$$\mathrm{a}_{02} = \frac{1-(-1)^{L}}{2}\frac{1 - \Omega_0}{\Omega_0 + 1}, \tag{23}$$

$$\mathrm{a}_{\mathrm{i}1} = \frac{4\mathrm{c}(\Omega_0\cos\Theta_{\mathrm{i}} - 1)}{1 - 2\Omega_0\cos\Theta_{\mathrm{i}} + \Omega_0^2}, \tag{24}$$

$$\mathrm{a}_{\mathrm{i}2} = \frac{2(2\mathrm{c}^2 + 1 - \Omega_0^2)}{1 - 2\Omega_0\cos\Theta_{\mathrm{i}} + \Omega_0^2}, \tag{25}$$

$$\mathrm{a}_{\mathrm{i}3} = -\frac{4\mathrm{c}(\Omega_0\cos\Theta_{\mathrm{i}} + 1)}{1 - 2\Omega_0\cos\Theta_{\mathrm{i}} + \Omega_0^2}, \tag{26}$$

and

$$\mathrm{a}_{\mathrm{i}4} = \frac{1 + 2\Omega_0\cos\Theta_{\mathrm{i}} + \Omega_0^2}{1 - 2\Omega_0\cos\Theta_{\mathrm{i}} + \Omega_0^2}, \tag{27}$$

in which Ω_0 changes from (20) to

$$\Omega_0 = -\frac{c - \cos\omega_{\mathrm{L}}}{\sin\omega_{\mathrm{L}}} = \frac{c - \cos\omega_{\mathrm{R}}}{\sin\omega_{\mathrm{R}}}, \tag{28}$$

where ω_{L} and ω_{R} are the left and right 3-dB cutoff frequencies, respectively. The parameter c is calculated using (29) and is written in terms of the center frequency ω_{c} of the desired band-pass filter as

$$c = \frac{\sin(\omega_{\mathrm{L}} + \omega_{\mathrm{R}})}{\sin\omega_{\mathrm{L}} + \sin\omega_{\mathrm{R}}} = \cos\omega_{\mathrm{c}} \tag{29}$$

The corresponding numerator is in the form $[(1+\mathrm{z}^{-1})(1-\mathrm{z}^{-1})]^{L}$.

The term $H^{'}(\mathrm{z})$ in (14) was determined from the known $A\,(\mathrm{z})$ and $B\,(\mathrm{z})$, and the gain factor G can be calculated using MPN. Since the maximum value of $H^{'}(\mathrm{z})$ must be normalized to 0 dB, the parameter G must be equal to the inverse of the maximum magnitude of the frequency response. Because of the monotonic magnitude responses of Butterworth filters, the resulting digital low-, high-, and band-pass filters obtain their maximum values at 0, π, and ω_{c}, respectively:

$$G_{\mathrm{LP}} = \frac{1}{\left|\mathrm{H}^{'}_{\mathrm{LP}}(\mathrm{e}^{\mathrm{j}\omega}|_{\omega=0})\right|}, \tag{30}$$

$$G_{\mathrm{HP}} = \frac{1}{\left|\mathrm{H}^{'}_{\mathrm{HP}}(\mathrm{e}^{\mathrm{j}\omega}|_{\omega=\pi})\right|} \tag{31}$$

and

$$G_{\mathrm{BP}} = \frac{1}{\left|\mathrm{H}^{'}_{\mathrm{BP}}(\mathrm{e}^{\mathrm{j}\omega}|_{\omega=\omega_{\mathrm{c}}})\right|}. \tag{32}$$

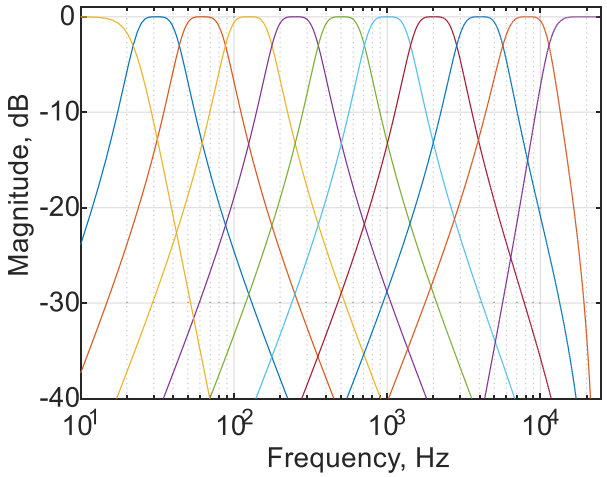

FIGURE 8. The proposed octave filterbank in the fourth order.

TABLE 1. Filter used in proposed equalizer.

Frequency band (Hz)	Center frequency (Hz)	Filter type
0 — 22	22 (cut-off frequency)	low-pass filter
22 — 44	31.5	octave filter
44 — 88	63	octave filter
88 — 177	125	octave filter
177 — 355	250	octave filter
355 — 710	500	octave filter
710 — 1420	1000	octave filter
1420 — 2840	2000	octave filter
2840 — 5680	4000	octave filter
5680 — 11360	8000	octave filter
11360 — 24000	11360 (cut-off frequency)	high-pass filter

The equalizer divides all frequency components into 11 frequency banks using the filters, which are listed in Table 1. The bandwidths of the band-pass filters are an octave. The cutoff frequencies of the low-pass filter and the high-pass filter depend on the left corner frequency of the first-octave filter and the right corner frequency of the last octave filter, respectively. The designed fourth-order octave band equalizer is shown in Fig. 8. In the conventional design, the first filter is a band-pass filter with a 16-Hz center frequency from 11 to 22 Hz, and the last filter is a band-pass filter with a 16-kHz center frequency from 11,360 to 22,720 Hz. In the proposed system, the low-pass filter and the high-pass filter are substituted for the band-pass filters to avoid the attenuation of extremely low (less than 11 Hz) and high (more than 22,720 Hz) frequency components.

The function of a fader is to vary the volume of the sound from each audio channel independently to balance the loudness of vocals and musical instruments. During a performance, it is usual to set the sound level around the midpoint of the scales. This leaves sufficient headroom to prevent clipping of sudden loud sounds while keeping enough loudness to mask any unwanted noise.

Panning controls set the balance of each input channel between the left and right output channels. The energy of the signal goes to the left channel if one turns the panning to the left and the sound energy goes to the right if the panning is turned to the right. When the panning is in the middle position, the listener perceives the same sound level from the left and right channels. Panning sets the location of the sound image in the stereo system. One normally places the vocal in the middle of the stage and spaces the musical instruments out slightly, which helps the audience to recognize individual sounds and provides a virtual acoustic space.

In the proposed system, the panning is not only for stereophonic but also for multichannel surround. If the user chooses to use a multichannel loudspeaker system, the sound image can be placed at any angle in a two-dimensional space. Taking Fig. 9 as an example, when the knob is turned to nine o'clock, the sound energy is equally distributed between the left front loudspeaker and the left rear loudspeaker, whereas the right front loudspeaker and the right rear loudspeaker are mute. The two-dimensional panning effect allows a more realistic and spatial sound field production than the conventional panning.

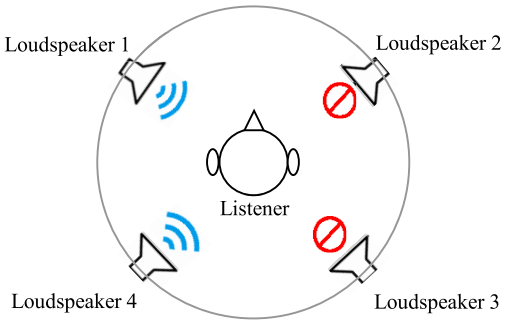

FIGURE 9. Panning for quadraphonic surround.

D. DENOISER

When singers perform on the street, background noise is likely to be picked up by their microphones, therefore degrading the sound quality. Although an equalizer can remove low- or high-frequency noises, the frequency components of the vocals in the same frequency band are also removed. In the proposed system, a special effect unit, a denoiser, is proposed to cancel the noise from busy streets.

The proposed denoiser is a modification and extension of the speech enhancement in the previous research [17]. In the original speech enhancement, the algorithm aimed to extract the low- or mid-frequency voice. The sound energy at frequencies of more than 5 kHz has normally been neglected in speech perception research [18]. However, the singers' vocals contain higher-pitch sound than speech. Therefore, some improvements were made for wide-frequency voice signal.

Because a general mobile device only contains one microphone input, a single-channel denoiser was developed. The first advantage of the proposed denoiser is data preprocessing. The data preprocessing can be done offline before the

performance and can be customized. That is, the user has to initially calibrate the denoiser. A piece of clean vocal $v[t]$ was recorded during vocalization in a quiet environment. The purpose of vocalization is to find the highest and lowest vocal notes that the user can produce. Then, the system designed a band-pass filter with the left cutoff frequency corresponding to the lowest vocal note and the right cutoff frequency corresponding to the highest vocal note. When the user performs in a noisy environment, the microphone signal is filtered by the customized band-pass filter.

After the band-pass filtering, the noise-reduction module illustrated in the previous work [17] was modified. The noisy vocal $m[t]$ was composed of $v[t]$ and the ambient sound $n[t]$, as shown in

$$m[t] = v[t] + n[t]. \tag{33}$$

A Wiener filter $f[t]$ was designed to extract the clean voice $v[t]$ from the noisy voice $m[t]$, so

$$\tilde{v}[t] = f * m[t], \tag{34}$$

where $\tilde{v}[t]$ is the estimated clean voice and $*$ is the convolution operator. The frequency domain expression of (34) is

$$\tilde{V}(\omega) = F(\omega) M(\omega). \tag{35}$$

According to the previous work [19],

$$F(\omega) = \frac{|M(\omega)|^2 - E[|N(\omega)|^2]}{|M(\omega)|^2}, \tag{36}$$

where E[·] is the ensemble average, and $N(\omega)$ and $M(\omega)$ are the frequency domain of the ambient sound $n[t]$ and noisy vocal $m[t]$. In the previous research [17], the noise level $N(\omega)$ was unknown and (36) was modified under the assumption that

$$E\left[|N(\omega)|^2\right] \approx (E[|M(\omega)|])^2. \tag{37}$$

However, (37) only happens when the ambient sound is stationary. In the proposed system, users record ambient sound $n[t]$ before the performance and the coefficients in (36) can be precisely computed.

Although the Wiener filter has been widely used in speech enhancement, the Kalman filter is more suitable for non-stationary signals. A Kalman filter was therefore added in the denoiser. The speech $v[t]$ was modeled as a p^{th} order autoregressive process [20]:

$$v[t] = -\sum_{i=1}^{p} l_i v[t\text{-}i] + u[t], \tag{38}$$

where l_i is the linear prediction coefficient (LPC) and $u[t]$ is the process noise. Equation (38) can be rewritten by the following matrix operation:

$$\begin{bmatrix} v[t-p+1] \\ v[t-p+2] \\ \vdots \\ v[t] \end{bmatrix} = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -l_p & -l_{p-1} & \cdots & -l_1 \end{bmatrix} \begin{bmatrix} v[t-p] \\ v[t-p+1] \\ \vdots \\ v[t-1] \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} u[t]. \tag{39}$$

A new notation was used to rewrite (39):

$$\boldsymbol{V}[t] = \emptyset \boldsymbol{V}[t-1] + \boldsymbol{O}u[t]. \tag{40}$$

The elements inside the matrix $\emptyset$ are LPCs and can be computed by an autoregression model. A matrix operation was used to represent (33):

$$m[t] = \boldsymbol{Z}\boldsymbol{V}[t] + n[t], \tag{41}$$

where $\boldsymbol{Z} = [00\ldots01]$. Assume the process noise $u[t]$ and the observation noise $n[t]$ are zero-mean signals and uncorrelated. The Kalman filter estimates the state vector $\boldsymbol{V}[t]$ by using the following recursive relations:

$$\hat{\boldsymbol{V}}[t|t] = \hat{\boldsymbol{V}}[t|t-1] + \boldsymbol{K}[t]\left(m[t] - \boldsymbol{Z}\hat{\boldsymbol{V}}[t|t-1]\right), \tag{42}$$

and

$$\hat{\boldsymbol{V}}[t|t-1] = \emptyset \hat{\boldsymbol{V}}[t-1|t-1], \tag{43}$$

where $\hat{\boldsymbol{V}}[t|t-1]$ is the *a priori* estimate of $\boldsymbol{V}[t]$, and $\hat{\boldsymbol{V}}[t|t]$ is the *a posteriori* estimate of $\boldsymbol{V}[t]$. $\boldsymbol{K}[t]$ is the Kalman coefficient and can be iteratively computed by

$$\boldsymbol{K}[t] = \left\{\boldsymbol{P}[t|t-1]\{\boldsymbol{Z}\}^T (\boldsymbol{Z}\boldsymbol{P}[t|t-1])\{\boldsymbol{Z}\}^T + \boldsymbol{R}\right\}^{-1}, \tag{44}$$

$$\boldsymbol{P}[t|t-1] = \emptyset \boldsymbol{P}[t-1|t-1]\{\emptyset\}^T + \boldsymbol{O}\boldsymbol{Q}\{\boldsymbol{O}\}^T, \tag{45}$$

and

$$\boldsymbol{P}[t|t] = (\boldsymbol{I} - \boldsymbol{K}[t]\boldsymbol{Z})\boldsymbol{P}[t|t-1], \tag{46}$$

where $\{\cdot\}^T$ denotes the transpose of a matrix; $\{\cdot\}^{-1}$ denotes the inverse; $\boldsymbol{R}$ and $\boldsymbol{Q}$ are the covariance matrices of the observation noise and process noise, respectively;$\boldsymbol{P}[t|t]$ is the error covariance matrix, and $\boldsymbol{I}$ is the identity matrix. The initial state vector $\hat{\boldsymbol{V}}[0|0]$ is the first p data points in $m[t]$.

Since the clean voice and the ambient sound were both recorded beforehand, it was possible to estimate which filter was appropriate for the current environment. The denoiser added up the clean voice $v[t]$ recorded during vocalization and the noise $n[t]$ recorded on the performance stage, as shown in (33). The Kalman filter and Wiener filter parallelly processed the $m[t]$. The estimated clean speech by the Kalman filter is denoted $\hat{v}[t]$, and that by the Wiener filter is denoted $\tilde{v}[t]$. The proposed system calculated the signal-to-noise ratios (SNRs) of $\hat{v}[t]$ and $\tilde{v}[t]$ and the appropriate filter could be selected.

As a result, the specification of the customized band-pass filter was designed by the observation of the user's vocal range. A dual-mode filter structure was implemented to overcome the various noise types in different environments.

III. AUDIO AMPLIFIER FOR MOBILE DEVICES

The input of an audio interface can be either line-level signals or microphone-level signals. The line-level signals normally come from electric instruments, such as an electric bass, electric guitar, or electric piano. The microphone signals have smaller voltages than the line-level signals and require preamplification. A typical mobile device only contains a 3.5-mm jack for microphone-level signals and therefore it was necessary to design an interface for electric instruments.

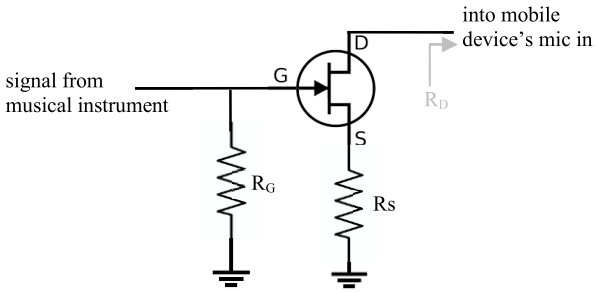

FIGURE 10. Junction field effect transistor (JFET) amplifier inside the proposed audio interface for mobile devices.

Because an amplifier in the interface should have high input resistance, a common-source junction field effect transistor (JFET) amplifier, as shown in Fig. 10, was applied. JFETs are most suitable for input stage designs where high input resistance is required [21]. Mobile devices typically supply a direct current voltage of approximately 2.5 V to the microphone, and therefore, the bias voltage can be applied to the drain. For impedance matching, it was necessary to design the resistor R_S, but the drain was connected to the mobile device, so the drain resistor R_D was hidden. In this case, the resistor R_S was designed by a trade-off. A 1.2-kΩ resistor is a reasonable value for most mobile devices. Because the input electric guitar signal is in the form of an alternating current, there are no coupling capacitors. The gate-source junction is reverse-biased, and there is only a very small leakage current, to possibly make the gate resistor very large. In practice, R_G is usually set to approximately 1 MΩ to avoid voltage drop [21]. The input of the interface is a 1/4-inch jack to plug in an electric guitar. The output is the four-conductor 3.5-mm plug connecting the interface to a mobile device.

The developed audio interface is shown in Fig. 11a, and its internal circuit is illustrated in Fig. 11b. The instrument signal was fed into the interface through a standard 1/4-inch mono plug and was amplified by the JFET amplifier. The amplified signal level was similar to the microphone signal level to transmit the treated instrument signal to the mobile device through a 3.5-mm jack. Then, the software effect units installed on the mobile device added special sound effects to the clean sound. An earphones plug was also created on the 3.5-mm jack, making it possible to monitor the sound effects by using external loudspeakers or headphones.

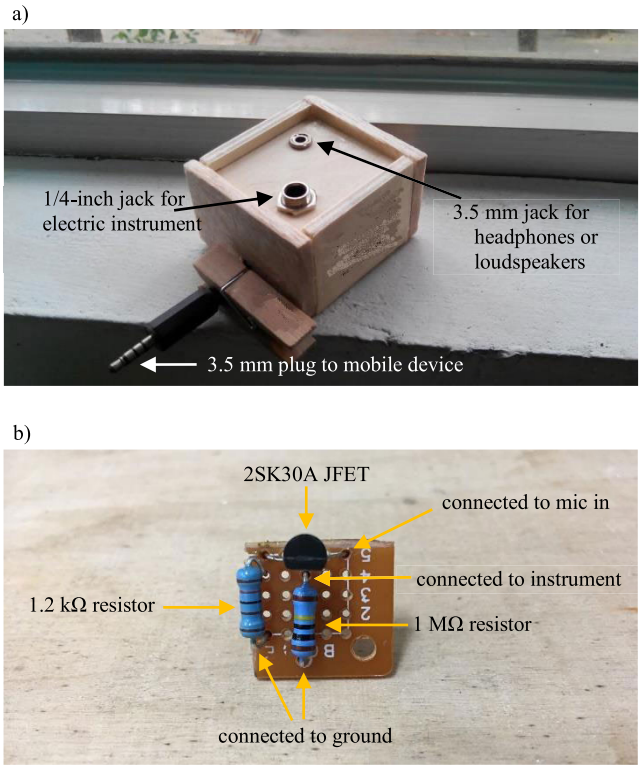

FIGURE 11. (a)Appearance and (b) internal circuit of the proposed audio interface.

IV. EXPERIMENTS AND DISCUSSION

The perceptions of special sound effects are subjective, and there is no golden model for comparison. Therefore, a subjective listening test was conducted for special sound effects in the proposed system. The questionnaire is shown in Table 2. Whereas Q1 to Q3 were designed to assess the envelopment of the delay effects, Q4 to Q6 were designed to assess the nature of the delay effects. Q7 to Q9 were designed to evaluate the practicability of the multichannel panning effect.

Because there were three delay effects to be evaluated, comparisons between the echo effect, reverberation effect, and room model effect were conducted. In the first question of the first question set, the subject was asked to choose the better envelopment effect between the echo effect and the reverberation effect. In the second question, the subject chose the better envelopment effect between the reverberation effect and room model effect. In the final question in this question set, the subject chose the better envelopment effect between the echo effect and room model effect. Therefore, the ranking in the aspect of envelopment quality was determined. The same procedure was conducted when evaluating the reality of delay effects from Q4 to Q6.

From Q7 to Q9, the conventional and the proposed panning techniques were compared. The conventional effect can only represent a one-dimensional sound field, whereas the proposed effect can produce a two-dimensional sound field. Therefore, the subject was asked to select the wider panning range in Q7. In Q8, whether there was a penalty for the wide panning range was checked. That is, the proposed effect may

TABLE 2. Questions and available answers for the subjective listening test.

No.	Question	Answer
Q1	Please listen to the two audio pieces, and find which one presents higher envelopment	A/B
Q2	Please listen to the two audio pieces, and find which one presents higher envelopment	A/B
Q3	Please listen to the two audio pieces, and find which one presents higher envelopment	A/B
Q4	Please listen to the two audio pieces, and find which one gets more realistic reverberation	A/B
Q5	Please listen to the two audio pieces, and find which one gets more realistic reverberation	A/B
Q6	Please listen to the two audio pieces, and find which one gets more realistic reverberation	A/B
Q7	Please try the two knobs, and find which one can place the sound image in wider range	A/B
Q8	Please try the two knobs, and find which one gets better audio quality	A/B
Q9	Please try the two knobs, and find which one presents higher auditory spatial resolution	A/B

or may not have a negative effect on the timbre of the sound. The resolutions of the sound image location were evaluated in Q9. Because a knob was used to present a 360° audio scene in the proposed method, the resolution might be degraded.

In the objective listening test, the focus was on testing two parts of the proposed system: the denoiser and the audio interface. In both parts, the clean voice or music served as the reference signal, and an objective computer-based algorithm, perceptual evaluation of audio quality (PEAQ), was applied. PEAQ is the International Telecommunications Union (ITU) standard for audio quality assessment [22], developed for the objective measurement of the perceived audio quality. Unlike traditional objective measurement methods, such as the SNR or the total harmonic distortion, PEAQ was designed based on neural-network training. The measurement scheme is shown in Fig. 12. The overall difference grade (ODG) was generated according to the comparison between the signal being tested and the reference signal. Normally, ODG has a value in the range of -4 to 0. As shown in Fig. 13, a more negative score indicates a more perceptible audio distortion. The ODG values provide information about the tolerance of audio distortion to evaluate the performances of the denoiser and the audio interface.

A. SUBJECTIVE LISTENING TEST FOR SPECIAL SOUND EFFECTS

In the subjective listening test, the aim was to find whether the multichannel surround improves the audio quality

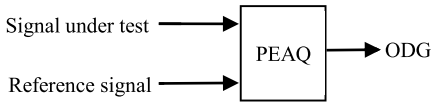

FIGURE 12. PEAQ measurement scheme.

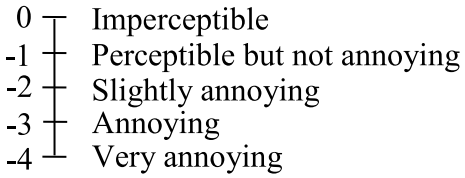

FIGURE 13. ITU-R five-grade scale.

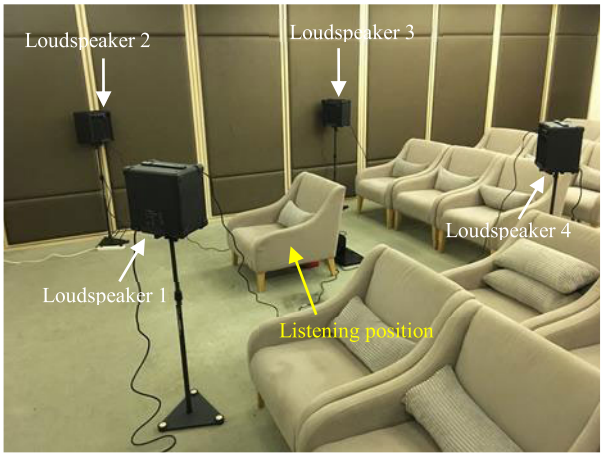

FIGURE 14. Loudspeaker array for subjective listening test.

and spatial listening perception. Four loudspeakers were mounted around the listener, as shown in Fig. 14, and the listener was asked to answer the questionnaire shown in Table 2. Stereophonic and quadraphonic sounds were played in random order. When playing stereophonic sound, loudspeakers 3 and 4 were muted. Only the proposed room model and the proposed panning effect simultaneously used four loudspeakers. During the listening test, the listener was not told that the sound field was produced by two or four loudspeakers.

There were 15 subjects involved in the listening test. There were 15 audio pieces, five of which were produced by a guitar, five of which were produced by a piano, and five of which were produced by a drum set. Three delay effects and two panning effects were tested. The best special-effect ranking is three points, while the lowest ranking is only one point. If there is an effect ranking in the middle, it is two points. The means and 95% confidence intervals of the answers were calculated and are shown in Fig. 15.

From Fig. 15, one can find that the proposed room model produced higher envelopment and more-realistic listening perception than the conventional delay effects. In a normal auditory space, the reverberations come from many directions, but the conventional delay effects can only convey the

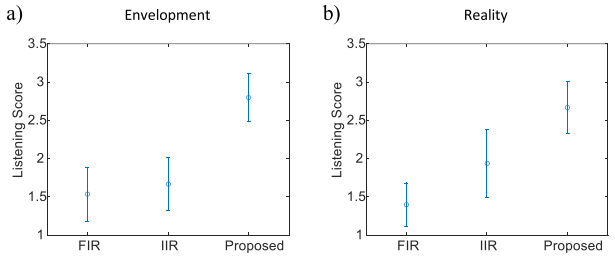

FIGURE 15. Statistical results of the listening perceptions in (a) envelopment and (b) reality in different delay effects.

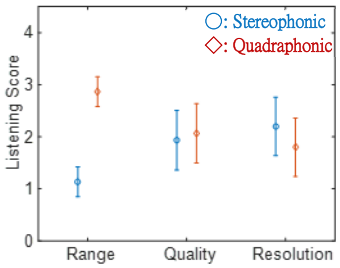

FIGURE 16. Statistical results of the listening perceptions in sound range, quality and resolution in different panning effects.

reflections in front of the listeners. The proposed room model used four loudspeakers to render a full two-dimensional auditory space to envelop the listeners in the sound field. The results also showed that the reverberation effect by the first-order IIR filter slightly outperformed the echo effect by the first-order finite impulse response filter.

Fig. 16 shows that the proposed two-dimensional panning effect performed better than the conventional panning effect in Q7. This is because the sound image cannot only be placed in front of the listener but can also be located behind the listener in the quadraphonic array. Q8 supported that the wide panning range did not degrade the timbral fidelity. However, the proposed panning effect presented lower auditory spatial resolution than the conventional one. The results of Q9 showed that the negative effect was perceptible but slight.

B. MEASUREMENTS OF DENOISER

The ambient sounds were recorded in five kinds of real environments: a basketball court, park, hotel lobby, exhibition room, and street. There were four professional singers — three males and one female — each of whom individually sang three pieces of a song in an anechoic chamber. Therefore, there were twelve different vocal pieces and five types of background noise, for a total of 60 combinations. The clean voice combined with noise was treated by the denoiser. The output of the denoiser was the signal being tested, and the clean voice was the reference signal in the PEAQ system.

The proposed denoiser was compared with two algorithms for noise reduction in the literature [17], [23]. The mean of overall ODGs is -3.685 in a previous study [17] and -3.695 in [23], which means the degradation of the processed sound is more than annoying. The mean of overall

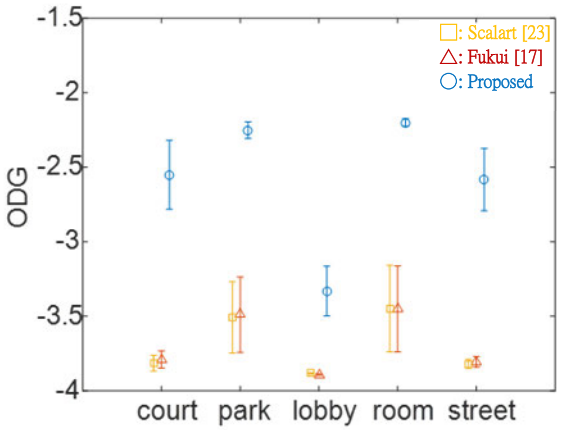

FIGURE 17. Evaluation of audio qualities in different noise reduction algorithms.

ODGs of the proposed denoiser is -2.583, which means the processed sound is between annoying and slightly annoying. The overall improvement of the proposed method is more than a grade, as shown in Fig. 13. Fig. 17 shows the mean ODGs and corresponding 95% confidence intervals in each environment, where one can see that all the denoisers had the worst performance in a hotel lobby. This might be because of too much human voice involved in the environment, violating the assumption of the algorithms that the sources should be independent of each other. The proposed denoiser had the best performance in the exhibition room. The ambient sound in the room mostly came from the air conditioners.

C. MEASUREMENTS OF AUDIO INTERFACE

The proposed audio amplifier was compared with a commercial product. Although SNR measurement has been widely used to evaluate the performance of analog amplifiers, different frequency components lead to different sensitivities and frequency resolutions for human hearing. The proposed interface was developed for audio signal processing only, and PEAQ was applied to the audio quality measurement according to psychoacoustics.

A personal computer (PC) and a smartphone were connected by using the developed interface. The PC was only used to emit the sound source. The voltage level of the music coming from the PC was line level, and the amplifier in the interface transferred the line level into the microphone level for the smartphone to record. The original music file was stored inside the PC, and the treated music file was generated by the smartphone. PEAQ measurement was used, as shown in Fig. 18, to evaluate the levels of distortion caused by the audio interfaces. There were three electric instruments — guitar, keyboard, and bass — each of which generated ten audio pieces. Fig. 19 shows the comparison between the proposed interface and the commercial product. According to the means of ODGs, the proposed architecture tended to obtain better audio quality performance. It was also found that the proposed audio interface outperformed the

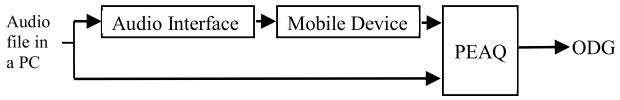

FIGURE 18. Audio interface measurement by PEAQ.

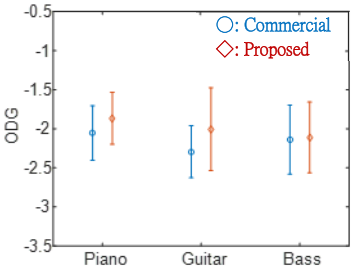

FIGURE 19. Comparison between the proposed audio interface and the commercial product.

TABLE 3. Comparison between interfaces at a life show.

Music Styles	ODGs for Commercial Product	ODGs for Proposed Interface
Folk Rock	-2.386	-2.351
Slow Soul	-2.375	-2.269
Rumba	-2.584	-2.270
Punk	-2.208	-2.183
Pop-Rap	-2.413	-2.236
Ballad	-2.308	-2.179
R&B	-2.246	-2.153
Country	-2.223	-2.150

commercial one when using electric guitars as input signals. This might be because the resistors in Fig. 10 were tuned by connecting an electric guitar to obtain the optimal audio quality.

A professional guitar player was also invited to conduct a live recording. The guitar player plugged the electric guitar into the audio interface and connected the interface to a smartphone. The audio recorder software was turned on in the smartphone to record the signal passing through the interface. Simultaneously, the clean guitar signal was recorded into a PC via a sound card. As a result, the audio file recorded in the PC was the reference signal, and the audio file in the smartphone was the signal being tested. The guitar player strummed four patterns corresponding to four music styles: folk rock, slow soul, rumba, and punk. The guitar player also played four different fingerstyle songs, including pop-rap, ballad, R&B, and country. The ODGs are shown in Table. 3, where it is found that the proposed system still outperformed the commercial one in all circumstances at the live show.

V. CONCLUSION

Professional audio equipment is too large and heavy for a single user to conduct a live performance. Therefore, in this study, an affordable mass-market system for ordinary consumers was proposed. Software was used to implement special sound effects with low-power audio interface hardware. Not only were several popular effect units implemented on mobile devices, but some novel sound effects that are rarely found on the market were also proposed. A light audio interface was developed for the connection between musical instruments and mobile devices without any external power supplies. For people who use the proposed system, the weight of the equipment can be greatly reduced. The compact size of the system enables users to bring additional multichannel surround to enhance audio quality.

The experiments were conducted mainly by using a typical smartphone. The results showed that the proposed effect units supporting multichannel surround sound enhanced spatial sound quality. The novel delay effect presented more-realistic reverberation than the conventional one. The two-dimensional panning effect provided a wider sound image than the conventional pan control. The experimental results also showed that the proposed denoiser achieved obvious improvement in voice quality enhancement. This is because users were asked to calibrate the denoiser for customization purposes. The denoiser also could adjust to the new environment by using a dual-mode filter structure. Because a conventional audio interface is designed for PCs and only has the Universal Serial Bus or FireWire communication protocol, an audio interface was developed introducing the electric audio signal into a mobile device. The comparison results showed that the proposed interface provided a clearer audio signal than the commercial product. In the future, all proposed units will be integrated and compared with the analog audio system by conducting additional subjective listening tests on real streets.

ACKNOWLEDGMENT

The author would like to thank the associate editor and anonymous reviewers for helpful suggestions. The author would also like to thank all the participants volunteering to join the subjective listening tests. This study has been in part supported by grant MOST 107-2221-E-305-010-MY2 from the Ministry of Science and Technology, Taiwan.

REFERENCES

[1] F. P. Ling, F. K. Khuen, and D. Radhakrishnan, "An audio processor card for special sound effects," in *Proc. IEEE Midwest Symp. Circuits Syst.*, Lansing, MI, USA, Aug. 2000, vol.2, pp. 730–733.
[2] K. Byun, Y.-S. Kwon, S. Park, and N.-W. Eum, "Digital audio effect system-on-a-chip based on embedded DSP core," *ETRI J.*, vol. 31, no. 6, pp. 732–740, Dec. 2009.
[3] L. Liu, J. Bär, F. Friedrich, J. Gutknecht, and S.-L. Tsao, "A low power configurable SoC for simulating delay-based audio effects," in *Proc. IEEE Int. Conf. Reconfigurable Comput. FPGAs*, Cancun, Mexico, Dec. 2012, pp. 1–6.
[4] T. Mohamadi, "Designing ISA card with easy interface," in *Proc. 9th East-West Design Test Symp.*, Sevastopol, Ukraine, Sep. 2011, pp. 372–376.
[5] P. Anghelescu, S. Angelescu, and S. Ionita, "Real-time audio effects with DSP algorithms and DirectSound," in *Proc. 6th Int. Conf. Electron., Comput. Artif. Intell.*, Bucharest, Romania, Oct. 2014, pp. 35–40.
[6] S. G. McGovern, "Fast image method for impulse response calculations of box-shaped rooms," *Appl. Acoustic*, vol. 70, no. 1, pp. 182–189, Jan. 2009.
[7] S.-N. Yao, "Headphone-based immersive audio for virtual reality headsets," *IEEE Trans. Consum. Electron.*, vol. 63, no. 3, pp. 300–308, Aug. 2017.

[8] S. R. Postrel, "Competing networks and proprietary standards: The case of quadraphonic sound," *J. Ind. Econ.*, vol. 39, no. 2, pp. 169–185, 1990.
[9] M. A. Gerzon, "Periphony: With-height sound reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2–10, 1973.
[10] S. K. Mitra and J. F. Kaiser, Eds., *Handbook for Digital Signal Processing*. New York, NY, USA: Wiley, 1993.
[11] S.-N. Yao, "Driver filter design for software-implemented loudspeaker crossovers," *Arch. Acoust.*, vol. 39, no. 4, pp. 591–597, 2014.
[12] S.-N. Yao, "Equalization in ambisonics," *Appl. Acoust.*, vol. 139, pp. 129–139, Oct. 2018.
[13] S. N. Yao, T. Collins, and P. Jančovič, "Hybrid method for designing digital Butterworth filters," *Comput. Elect. Eng.*, vol. 38, no. 4, pp. 811–818, Jul. 2012.
[14] S. J. Orfanidis, *Introduction to Signal Processing*. New Jersey, NJ, USA: Prentice-Hall, 1995.
[15] C.-C. Tseng and S.-C. Pei, "Stable IIR notch filter design with optimal pole placement," *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 2673–2681, Nov. 2001.
[16] T. van Waterschoot and M. Moonen, "A pole-zero placement technique for designing second-order IIR parametric equalizer filters," *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, no. 8, pp. 2561–2565, Nov. 2007.
[17] M. Fukui, S. Shimauchi, Y. Hioka, A. Nakagawa, and Y. Haneda, "Acoustic echo and noise canceller for personal hands-free video IP phone," *IEEE Trans. Consum. Electron.*, vol. 62, no. 4, pp. 454–462, Nov. 2016.
[18] B. B. Monson, E. J. Hunter, A. J. Lotto, and B. H. Story, "The perceptual significance of high-frequency energy in the human voice," *Front Psychol.*, vol. 5, no. 587, pp. 1–11, Jun. 2014.
[19] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.
[20] K. K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Dallas, TX, USA, Apr. 1987, pp. 177–180.
[21] J. D. Denton, *Electronics for Guitarists*. New York, NY, USA: Springer, 2011.
[22] *Multichannel Stereophonic Sound System With and Without Accompanying Picture*, document Rec. BS.775-1, ITU-R, International Telecommunication Union, Geneva, Switzerland, Nov. 1994.
[23] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Atlanta, GA, USA, May 1996, pp. 629–632.

SHU-NUNG YAO (S'11–M'13) received the B.Sc. degree in electrical engineering from the National Taipei University of Technology, the M.Sc. degree in electrical engineering from National Cheng Kung University, and the Ph.D. degree in electronic, electrical and systems engineering from the University of Birmingham, U.K. He was an IC Design Engineer with Silicon Integrated Systems Corporation, from 2007 to 2008. From 2009 to 2010, he was with the Keelung Customs Office. In 2015, he founded ODOtek Company Ltd. Since 2016, he has been an Assistant Professor with National Taipei University, Taiwan. He was recognized as an Outstanding Reviewer of Computers and Electrical Engineering, in 2015 and 2017. He is a member of the AES and IEICE. His research has been centered around the field of signal processing, usually for acoustics-related applications.