99.4% Peak Audio Signal Recovery Rate and Ultra-Low 0.32dB Matching Error with 10Hz High Resolution Filter Fitting Wearable Aided Speech Compensation System

Yen-Ting Lin[1], Shin-Chi Lai[2], Shin-Hao Chen[1], Shen-Yu Peng[1], Ke-Horng Chen[1], Sheng Kang[3], Kevin Cheng[3], Ying-Hsi Lin[4], Chen-Chih Huang[4], and Chao-Cheng Lee[4]

[1] Institute of Electrical Engineering, National Chiao Tung University, Taiwan, [2]Dept. of Computer Science and Information Engineering, Nan Hua University, [3]Anpec Electronics Corporation, [4]Realtek Semiconductor Corporation, Hsinchu, Taiwan

***Abstract*- to get high sound quality, the proposed adaptive speech compensation (ASC) system fits psychoacoustic curves for individual needs. Moreover, characteristics of the speakers are considered by real-time speaker impedance estimation (RT-SIE) technique. Hence, original signal can be restored for every user with any speakers. ASC system has 99.4% peak original signal recovery rate by predicting and controlling output signal precisely. And the built-in 10Hz high resolution acoustic filter fitting scheme (AFFS) has less than 0.32dB matching error with an over 80dB hearing loss.**

***Keywords*—Psychoacoustic model, speaker impedance, class D audio amplifier, digital signal processing.**

I. Introduction

Of the 26.7 million people over age 50 with a hearing impairment, only one in seven, or 17%, use a hearing aid. Most of the cell phones operate in frequencies which interfere with hearing-aid devices. This paper places the focus on the sound quality improvement for those devices coupled with speakers, such as cell phones and audio acoustic systems, for hearing impaired people without wearing the hearing aids.

This work presents the cooperation of real-time speaker impedance estimation (RT-SIE) and audio signal processing algorithm to get high quality sound demanded by portable devices such as cell phones or wearable aided speakers. The paper presents an ASC system with the RT-SIE technique and the built-in AFFS. The proposed ASC system can significantly benefit to at least, but not limited to, those people with minor hearing loss still use the cell phones and enjoy the music well. As illustrated in Fig. 1, speech voice and sound quality greatly suffer from individual psychoacoustic response. Even if individual hearing response can be derived, high quality sound cannot be played by every speaker because the characteristics of different speakers vary over a wide range. The ASC system provides several speech compensation modes to satisfy the individual needs. By fitting psychoacoustic curves, original audio signal can be restored for every user. With auxiliary RT-SIE technique, real-time characteristics of the speaker can be measured. Perfect fitting for both psychoacoustic response and RT-SIE of the speaker guarantee high sound quality [1, 2].

As shown in Fig. 2, hearing threshold of human ears is frequency dependent. By fitting psychoacoustic curves, the conversion from electrical power to sounds can be more efficient. There are two features in a human ear model. Firstly, the absolute threshold of hearing (TOH) curve, which varies along with the frequency from 20Hz to 20kHz as illustrated in (1).

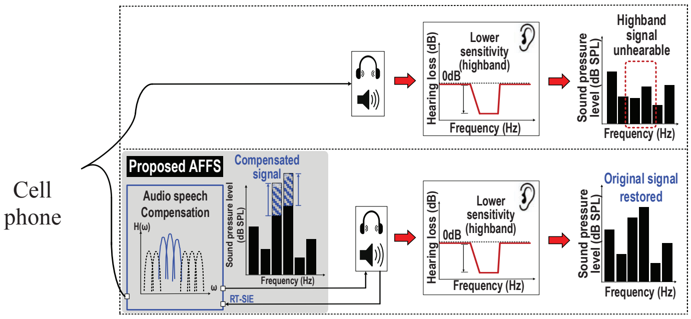
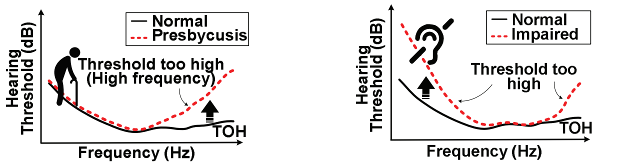

Cell phone
Lower sensitivity (highband)
Hearing loss (dB)
0dB
Frequency (Hz)
Highband signal unhearable
Sound pressure level (dB SPL)
Frequency (Hz)
Proposed AFFS
Audio speech Compensation
H(ω)
ω
Compensated signal
Sound pressure level (dB SPL)
Frequency (Hz)
RT-SIE
Lower sensitivity (highband)
Hearing loss (dB)
0dB
Frequency (Hz)
Original signal restored
Sound pressure level (dB SPL)
Frequency (Hz)

Fig. 1. Proposed speech compensation diagram on a cell phone application.



(a) (b)

Fig. 2. Hearing threshold of human ears. (a) presbycusis and (b) impaired person.

$$TOH(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5e^{\left(\frac{f}{1000}-3.3\right)^2} + 10^{-3}\left(\frac{f}{1000}\right)^4 \text{ (dB)} \qquad (1)$$

The absolute value is *TOH*(*f*) = 0 when *f* = 2kHz. The threshold curve changes with different people and different ages. Along with the growth of the age, the threshold will become higher and the hearing frequency range will decrease as shown in Fig. 2(a). Moreover, TOH may be much worse in a hearing impaired person as illustrated in Fig. 2(b). Moreover, the second feature is the masking effect which includes spectral masking in frequency domain and temporal masking in time domain. Spectral masking indicates that, it interferes all the sounds from neighbor frequency range when a loud single frequency sound is played. Actually, every single tone can introduce its new masking curve to replace TOH in the nearby region. For example, the original TOH is replaced by a new masking threshold introduced by a 260Hz single tone. The threshold has been greatly increased around 260Hz. As a result, the second tone at 150Hz is masked by the new TOH and cannot be heard anymore.

Speech frequency band from 20Hz to 8kHz can be divided into 18 critical bands for a trial in this paper. In lower frequency bands, the bandwidth of one critical band is smaller than 100Hz. On the other hand, the bandwidth of one critical band is larger than 1kHz in higher frequency bands. Therefore, human ears can distinguish sounds with higher resolution at low frequencies.

978-1-4673-7191-9/15/$31.00 ©2015 IEEE

In general, two tones within one critical band cannot be distinguished during normal condition. For a given central frequency f_c, the bandwidth of critical band is derived in (2).

$$df(Hz) = 25 + 75 \times \left[1 + 1.4 \times f_c^{\,2}(kHz)\right]^{0.69} \qquad (2)$$

Since bandwidth of critical bands changes along with frequency. A new frequency unit, Bark, has been defined for getting a similar bandwidth in every critical bandwidth as illustrated in (3).

$$b\ (Bark) = 13 \cdot \tan^{-1}(0.76f) + 35 \cdot \tan^{-1}\left(\frac{f^2}{56.25}\right) \qquad (3)$$

where b is the number of critical bands in Bark unit and the unit of f is kHz. Bandwidth of every critical band is almost the same in Bark unit. On the contrary, (4) shows the transformation from b (Bark) to f (kHz).

$$f\ (kHz) = \left(\frac{e^{0.219 \times b}}{352} + 0.1\right) \times b - 0.032 \times e^{-0.15 \times (b-5)^2} \qquad (4)$$

To sum up, the features of masking effect is as follow: Lower frequency sounds can easily mask the relative higher frequency sounds. Masking effect from higher frequency sounds to lower frequency sounds is relatively poor. The range of masking region increases with the volume of sound. In this paper, critical band analysis is adopted to power saving to enhance the experience of hearing in an audio system.

II. Proposed Adaptive Speaker Compensation

The proposed ASC system can in real time achieve speaker impedance estimation to protect the speaker in Fig. 3. It is also based on well-known psycho-acoustic model. However, it might reduce the quality of sound a little bit. To high resolution, *i.e.* 10 Hz, acoustic filter fitting scheme and WOLA FFT filterbank are both added to keep a good sound quality according to subjective blind test. Fig. 3 shows the signal flow block diagram of proposed ASC system, which consists of one 2.5W class-D amplifier supplied by 5V and with DC resistance 8Ω, one digital-to-analog converter (DAC), two analog-to-digital converters (ADCs), one DSP core, data buffering memories and coefficient memories. The speech compensation system comprises four sub-modules. First of all, a reception module, Fast Fourier transform (FFT), is placed at input stage for transforming time-domain signals to frequency domain V_0 to V_{M-1} where M is the FFT-size. Moreover, the speaker impedance varies with inductor current and voltage-drop across the resistance at different frequencies. The RT-SIE module detects load current, I_L [4] and voltage, V_L to calculate real-time speaker impedance to adjust output power with higher accuracy. Thirdly, the prediction module predetermines an initial speech information to generate a power prediction information according to the audio spectrum information $V_k(n)$ and real-time impedance information $Z_f(n)$ which controls transfer function $h(0)$ to $h(M-1)$ in the next stage. The last one, a power adjustment module which contains various speech compensation coefficients for mode selection, links the prediction module and the reception module. Correspondingly, the power adjustment module outputs adjustment audio signals to the amplification stage according to power prediction. Hence, audio digital signals and control signals perform a broadcast operation of the loading speaker or earpieces.

A. Real-time speaker impedance estimation (RT-SIE) technique

For audio processing, since speaker impedance is frequency dependent as shown in Fig. 4(a). A commercial 8 Ω speaker, for example, has DC-resistance R_e 8 Ω at 0 Hz. The typical resonant frequency f_C is about 100 Hz which has the highest resistance about tens of Ohm. At higher frequencies, the speaker impedance increases along with frequency. As shown in Fig. 4(b), the impedance of a loudspeaker can be derived from (5) which is the measurement condition.

$$Z_{SPK} = R\frac{V_{SPK}}{V_s\left(1 - \frac{V_{SPK}}{V_s}\right)} \qquad (5)$$

DC resistance R_e and typical resonant frequency f_C can be directly identified by fitting the low frequency asymptote of the impedance value and the center frequency of the resonance peak. And the impedance Z_{SPK} within resonance frequency area is illustrated in (6) where the corresponding frequencies are identified as f_l and f_h.

$$Z_{SPK} = \sqrt{R_e R_c} \qquad (6)$$

R_c is the peak impedance of a resonance frequency. The low frequency behavior of a loudspeaker can be modeled with Thiele/Small parameters [3]. As shown in Fig. 4(c), even with the same input signal, speaker output signal are different from each other since the different impedance responses. In the ASC system, the DSP core receives the measurement signal and calculates whole impedance curve through Fast Fourier Transform (FFT) analyzer.

B. Weighted-overlap-add (WOLA) FFT-based filterbank

In this paper, total 12 processes are applied to analyze and synthesize the class-0 (the highest class) ANSI S1.11 digital filterbanks as shown in Fig. 5(a). The input signals are firstly sent into frame-based buffer and divided into M-point per frame. After M-point FFT operation, the frequency-domain signals are temporarily stored in another data buffer, and are further divided into the corresponding 18 subbands according to the specification of ANSI S1.11. Channel switching block can allow user deciding the desired band to generate the output signals. Then, each operation result of channel switching is multiplied by the corresponding coefficients to meet the specifications of transition-band and pass-band filtering for each subband. After subband filtering, the gain control block was utilized to process the auditory compensation through a prescription procedure. For 18-subbands, M-point IFFT operation is required for data recovering time-domain signal. It is worthy noticed that the group delay between each subband is the same. To avoid the boundary discontinuity between frames and to cancel a crack noise, hanning window function is applied in the proposed WOLA FFT-based filterbank. After smoothing the recovered signals by the post window, a constant-multiplication procedure is applied to scale down or up the amplitude. Finally, an overlap-and-addition method is employed to reconstruct the synthesis signals.

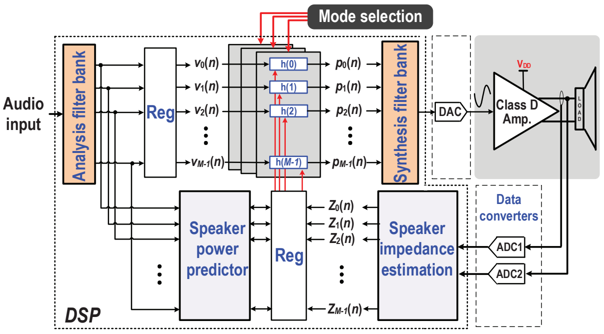
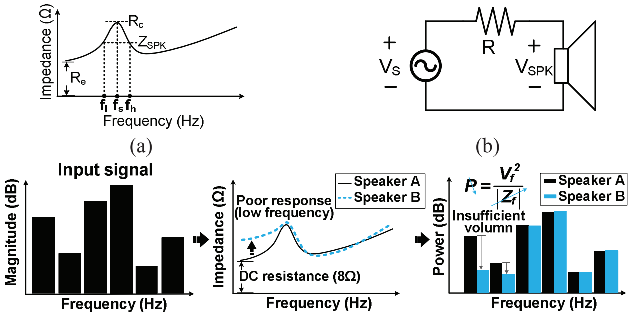


Fig. 3. Signal flow block diagram of proposed ASC system.



(a)

(b)

(c)

Fig. 4. (a) Speaker impedance characteristic. (b) Measurement setup. (c) Output power of different speaker impedance.

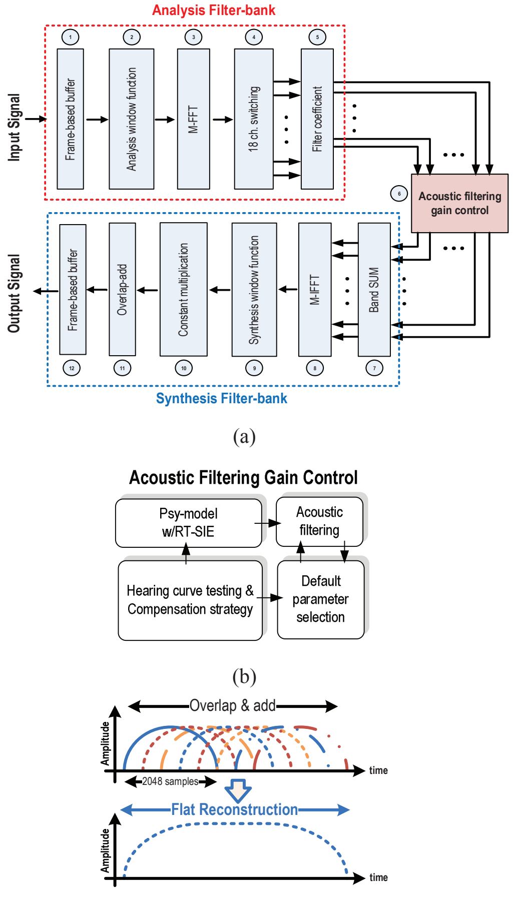
Fig. 5(a) demonstrates the proposed WOLA FFT-based filterbank where the FFT transform length is 2048. For a 10Hz frequency resolution per bin, the sampling rate is set to 20.48k Hz. Fig. 5(b) is the gain control block. In general, it can also be treated as an audio equalizer. Then, in Fig. 5(c), 75% overlapping and hanning window are used to improve the effects of spectrum leakages and Gibbs phenomenon [5].

C. High resolution acoustic filter fitting scheme (AFFS)

After sub-band filtering, the gain control block is utilized to proceed the auditory compensation by acoustic filter fitting scheme. To effectively recover the unplugged hearing response for users, a hearing pre-test is used to evaluate the sound-pressure-level (SPL) based treading curve for hearing sense. To avoid uncomfortable listening, half the amount of hearing loss for the pre-test is recommended for compensation in Fig. 5(b), and then makes a sound balance for acoustic equalization after combining with the absolute hearing threshold provided by psychoacoustic model. It is worth to notice that the group delay between each sub-band is the same. After smoothing the recovered signal through post-window, a constant multiplication procedure to scale down or up the amplitude is also needed. Finally, an overlap-and-addition method is employed to reconstruct the synthesis signal.

III. Matching Results

Fig. 6 illustrates matching results for four experimental subjects. For a given pre-test audiogram, 18 frequency bands ranging from 250Hz to 8KHz are considered to apply for speech compensation. Subject A hearing ability is slightly insensitive at high frequency bands. Subject B has a slight hearing insensitivity from 25dB to 33dB in full bands. Subject C has a moderate hearing loss at low frequency bands. And subject D has a severe hearing loss over 90dB. The constructed filters for each subject have little matching errors not greater than 0.32dB. By matching results, speech compensation rate can achieve 99.4% peak recovery rate. Audio fidelity and speech communication quality can be restored for every user in any wearable aided speakers.

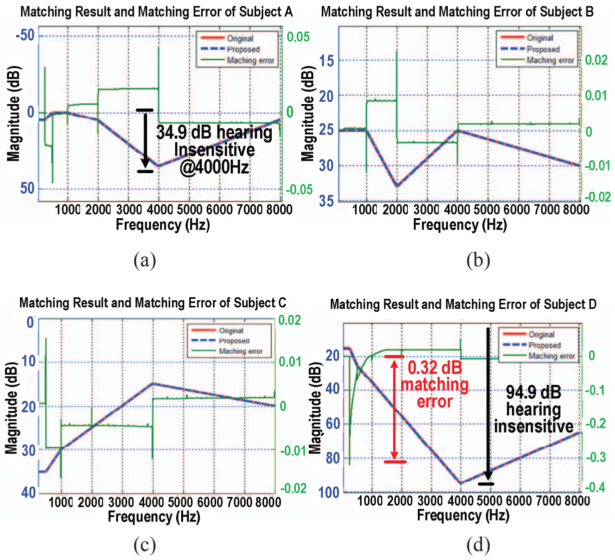


(a)

(b)

(c)

Fig. 5. (a) WOLA FFT-based filterbank, (b) gain control block and (c) overlapping and hanning window.



(a)

(b)

(c)

(d)

Fig. 6. Matching results for subject (a) A, (b) B, (c) C and (d) D.

IV. EXPERIMENTAL RESULTS

Experimental results are shown in Fig. 7. For a given piece of music, volume is too low at around 4KHz frequency band for subject A. By the ASC system, audio signal is compensated according to the fitting curve in Fig. 6(a). As a result, the piece of music heard by subject A can be restored by the ASC.

Fig. 8 shows a subjective blind test listening evaluation of the three different solutions: traditional operation, speaker output with RT-SIE and output with both RT-SIE and AFFS. There are two kinds of testing data: music and speech. Tests were done by 3 groups of listeners: youngsters, presbycusis and hearing impaired people. There are 20 listeners in each group. The listeners gave grades to the audio quality from 0 (bad) to 5 (good). At first, when audio output is music, the sound quality is slightly higher with RT-SIE in every group. And scores are significantly improved when the solution with both RT-SIE and AFFS. Secondly, in speech test, there is little benefit gained from RT-SIE. However, presbycusis and hearing impaired people have about 1.5 points notably progress when AFFS has been added. The comparison table illustrates -85dB most stop band attenuation and RT-SIE application by least computation effort.

The proposed FFT-based filterbanks design can achieve the highest level of ANSI S1.11 specification with the advantage of zero band group delay mismatch due to the nature of FFT property. In computation, it only requires 141 multiplications, 409 additions, and 1536 coefficients (1024 for hanning window and 512 for twiddle factors of FFT). Fig. 9 shows the experimental setup, silicon implementation and gate count distribution. It was realized by the TSMC 0.18µm 1P6M CMOS technology. The core size is 1.4×1.4 mm^2 and the power consumption is 3.01mW at 3.32MHz clock rate with 1V operating voltage.

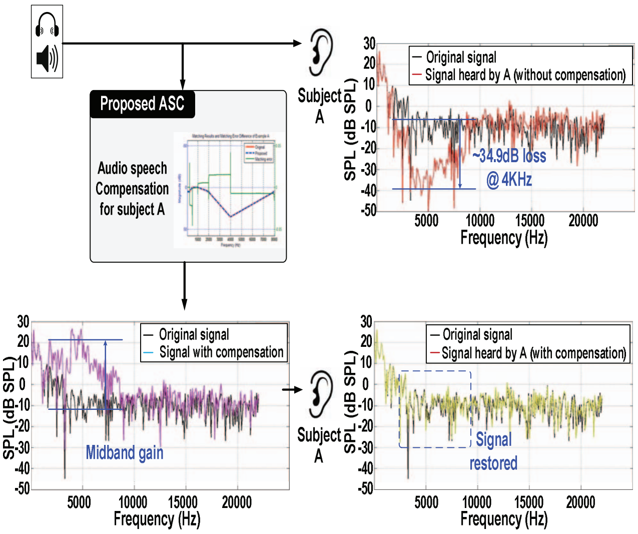
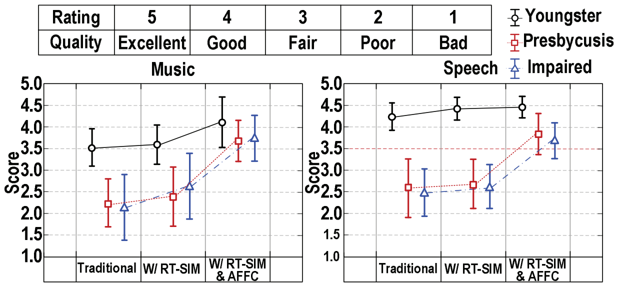


Fig. 7. Experimental result of subject A compensation. The upper part is the signal heard by A without compensation, the lower part is the signal with pre-compensation. The original signal can be restored.



Rating	5	4	3	2	1
Quality	Excellent	Good	Fair	Poor	Bad

Fig. 8. Subjective blind test listening evaluation and table of corresponding scores.

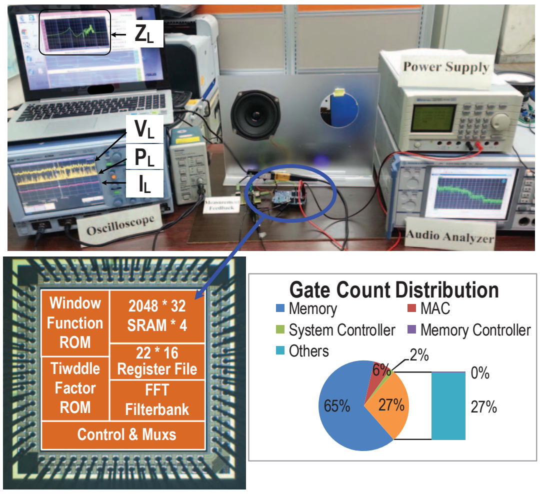


Fig. 9. Experimental setup, chip micrograph and gate count distribution.

Table I: Comparison of various filter bank designs.

Method	[1]	[2]	[3]	[5]	Proposed
Class-2	Satisfied	Satisfied	Quasi	N/A	Class 0
Structure	Multirate	Parallel	Multirate	Parallel	FFT-base
Sampling Rate (kHz)	24	24	24	24	20
Band	18	18	18	16	18
Mpy/S	140	3270	235	0	141
Add/S	278	6545	495	6720	409
# Coeff.	91	3270	506	880	1536
Channel group delay mismatch	77.5ms	9.3ms	9.5ms	3.4ms	0
RT-SIE	No	No	No	No	Yes
Stop band atten. (dB)	-60dB	-60dB	≒-60dB	-60dB	-85dB

References

[1] Y.-T. Kuo, T.-J. Lin, Y.-T. Li, and C.-W. Liu, "Design and implementation of low-power ANSI S1.11 filter bank for digital hearing Aids," *IEEE Tran. Circuits Syst. I, Reg. Papers*, vol. 7, pp. 1684-1696, 2010.

[2] C.-W. Liu, K.-C. Chang, M.-H. Chuang and C.-H. Lin, "10-ms 18-Band Quasi-ANSI S1.11 1/3-Octave Filter Bank for Digital Hearing Aids," *IEEE Tran. Circuits Syst. I, Reg. Papers*, vol. 60, no. 3, pp. 638-649, 2013.

[3] R H Small. Direct-Radiator Loudspeaker System Analysis. J. Audio Eng. Soc., 1972; 20(5): 383–395

[4] A. Nagari, E. Allier, F. Amiard V. Binet and C. Fraisse, "An 8Ω 2.5 W 1%-THD 104 dBA Dynamic-Range Class-D Audio Amplifier With Ultra-Low EMI System and Current Sensing for Speaker Protection," *IEEE Journal of Solid-State Circuits (JSSC)*, pp. 3068-3080, Dec. 2012.

[5] K.-S. Chong, B.-H. Gwee and J.-S. Chang, "A 16-channel low-power nonuniform spaced filter bank core for digital hearing aid," *IEEE Tran. Circuits Syst. I, Reg. Papers*, pp.853 -857, 2006.