Phoneme-Based Self Hearing Assessment on a Smartphone

Jong Min Choi, Junil Sohn, Yunseo Ku, Dongwook Kim, and Junghak Lee

***Abstract*—Phonemes provide an interesting alternative to pure tones in hearing tests. We propose a new smartphone-based method for self-hearing assessment using the four Korean phonemes which are similar to the English phonemes /a/, /i/, /sh/, and /s/. We conducted tests on 15 subjects diagnosed with mild to severe hearing loss and estimated their conventional pure-tone hearing thresholds from their phoneme hearing thresholds using regression analysis. The phoneme-based self-hearing assessment was found to be sufficiently reliable in estimating the hearing thresholds of hearing-impaired subjects. The difference between the hearing thresholds obtained through conventional pure-tone audiometry and those obtained using our method was 5.6 dB HL on average. The proposed hearing assessment was able to significantly reduce the mean test time compared to conventional pure-tone audiometry.**

***Index Terms*—Phoneme, self hearing assessment, smartphone.**

I. Introduction

SMARTPHONE is not only a mobile phone but also a hand-held computer with Internet connectivity. New researches of e-health applications using a smartphone as an e-health gateway or mobile healthcare device have been conducted [1]–[3]. Operating platforms for a smartphone such as Apple iOS and Google Android provide an excellent environment for e-health application and it is not becoming a difficult work to use this environment as it was several years ago.

Smartphones are becoming increasingly powerful, with modern audio codec chips providing smartphone audio of high quality. This makes it possible to conduct hearing tests like pure-tone audiometry (PTA) on a smartphone. Although commercial audiometers provide a dynamic range of more than 100 dB in air-conduction audiometry, which cannot be matched by current audio codec chips, the possibility of implementing hearing tests on a smartphone is already widely accepted. The existing literature proposes PTA on personal computers without Internet [4], [5] and with Internet [6], [7] and mobile phones [8]. The previous studies have tried to show that the accuracy of estimating hearing thresholds was comparable to that of PTA.

Manuscript received April 10, 2012; revised August 25, 2012 and December 6, 2012; accepted December 21, 2012. Date of publication January 9, 2013; date of current version May 1, 2013. This work was supported by the Strategic Technology Development Program of the Ministry of Knowledge Economy under Grant 10031741.

J. M. Choi, J. Sohn, Y. Ku, and D. Kim are with Future IT Research Center, Samsung Advanced Institute of Technology, Yongin 446-712, Korea (e-mail: jm1976.choi@samsung.com; ji.sohn@samsung.com; yunseo.ku@samsung.com; steve7.kim@samsung.com).

J. Lee is with the Department of Audiology, Hallym University of Graduate Studies, Seoul 135-280, Korea (e-mail: leejh@hallym.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2013.2238549

PTA is usually conducted in a sound-proof booth by an audiologist. The operator chooses testing frequencies in the following order: 1000, 2000, 4000, and 8000 Hz and then 500 and 250 Hz. There are three strategies to vary sound levels: the descending method, the ascending method, and the mixed method [9], [10]. In the mixed method, an audiologist initially plays a pure tone of 1000 Hz at a comfortable sound level. If a subject can hear this tone, the sound level is reduced by 10 dB HL; otherwise, the sound level is increased by 5 dB HL. The sound level at which the subject can hear a tone two times out of three is deemed to be the hearing threshold. The procedure is repeated at the other frequencies. PTA needs a relatively long testing time because subjects need to repeatedly respond to many stimuli at different frequencies and sound levels.

Speech audiometry is a form of audiometry that uses words instead of pure tones. Clinically, speech intelligibility is the most important aspect for the patients. There have been some reports about the difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram [11], [12]. Therefore, the method using complex tones such as speech should be performed in order to more accurately assess hearing ability of the patients [13]. We propose a phoneme-based approach that can be considered a variant of speech audiometry. We use regression-based methods to convert phoneme-estimated hearing thresholds into conventional audiogram thresholds.

We proposed a phoneme-based self-hearing assessment (PhoSHA) in our previous paper [14] and now implement it on a smartphone running the Android operating system. In addition, we verified our method with test samples which were not included when we calculated regression formulas unlike our previous study. The PhoSHA method provides an alternative way to test hearing thresholds with sufficient accuracy and reduced test time. Like the results of conventional audiograms, the results of this method can be used to fit hearing aids.

II. Method

Pure tones were replaced by four Korean phonemes in our proposed self-hearing assessment. Phonemes have multiple formants and, compared to tones, a wider spectral energy distribution and much more information content. We chose the four Korean phonemes which are similar to the English phonemes /a/, /i/, /sh/, and /s/. The phonemes were recorded by a female announcer. The vowel /a/ has three formants at 710, 1100, and 2640 Hz, /i/ has three formants at 400, 1900, and 2550 Hz, and the consonants of /s/ and /sh/ have distinguishable spectral characteristics from 3000 to 4000 Hz and from 2500 to 4000 Hz, respectively.

PhoSHA was implemented on a smartphone (SHW-M250S, Samsung Electronics, Suwon, South Korea) running the

Fig. 1. PhoSHA application on the Samsung smartphone.

Android operating system (Google Inc., Mountain View, CA, USA); it can also be calibrated for other Android smartphones. The earphones bundled with the smartphone were used for our experiments. Calibration was done for both left and right earpieces of the earphone. Four phonemes were all calibrated at 70 dB SPL with broad band setting of sound level meter (Bruel & Kjaer Type 2250). Fig. 1 shows the PhoSHA application on the Samsung smartphone.

The application plays the phonemes within a 30–85 dB SPL range. The level is automatically increased or decreased by 5 dB SPL according to the response of the subject. To begin with, /a/ is played at 50 dB SPL. If the subject responds to this sound, it is played at a lower level. If the subject does not respond, it is played at a higher level. After determining the threshold for /a/, the application moves on to /i/, /sh/, and /s/. We recorded the time taken to perform the test for each subject in addition to their hearing threshold. To reduce the test time, we used the threshold for /a/ (previous phoneme) as an initial sound level of /i/ (current phoneme), and the same procedures were repeated for the next phonemes. Practically, the initial sound level for /a/ was 35 dB SPL and the initial sound level for the next phoneme is set to 5 dB SPL lower than the threshold level of the previous phoneme. For example, if a subject could hear /i/ at 50 dB SPL, the initial sound level of /sh/ is set to 45 dB SPL to reduce test time.

We recruited 15 subjects diagnosed with mild to severe hearing loss and tested both ears of the subject. Three of these ears were normal hearing; 27 ears were tested. Two of these ears were excluded because of experimental problems; we finally tested 25. The number of left ears was 13 and the number of right ears was 12. The age range was 20–78 years, with a mean age of 69.9 years and a standard deviation of 12.2 years. We also recruited 17 normal subjects and analyzed 31 ears in this group. The age range of these subjects was 24–41 years, with a mean age of 28.9 years and a standard deviation of 5.0 years.

The phoneme-based hearing thresholds were converted into audiogram hearing thresholds using the regression formulas in Table I. Th_a, Th_i, Th_{sh}, and Th_s denote the measured hearing threshold levels (in dB SPL) of the test phonemes /a/, /i/, /sh/, and /s/, respectively. eTh_{250} to eTh_{8000} denote the estimated hearing thresholds at various frequencies. The regression coefficients in the formulas shown in Table I were derived using the measurements for all 25 ears and were calculated via least-squares. In accordance with convention, we rounded the values to the nearest 5 dB.

TABLE I
REGRESSION FORMULAS AND STATISTICAL PARAMETERS AT SIX FREQUENCIES

Frequency	Regression Formula	r^2	F	p-value
250	$eTh_{250} = 0.6Th_a + 0.14Th_i - 0.38Th_{sh} + 0.39Th_s - 4.81$	0.49	4.9	$p < 0.01$
500	$eTh_{500} = 0.89Th_a - 0.11Th_i - 0.59Th_{sh} + 0.66Th_s - 4.33$	0.59	7.3	$p < 0.001$
1000	$eTh_{1000} = 1.38Th_a - 0.27Th_i - 0.03Th_{sh} + 0.08Th_s - 18.9$	0.75	15.0	$p < 0.001$
2000	$eTh_{2000} = 0.45Th_a - 0.01Th_i + 0.84Th_{sh} - 0.15Th_s - 17.26$	0.78	17.4	$p < 0.001$
4000	$eTh_{4000} = 0.05Th_a - 0.26Th_i + 0.83Th_{sh} + 0.25Th_s - 0.8$	0.74	13.9	$p < 0.001$
8000	$eTh_{8000} = -0.22Th_a - 0.06Th_i + 0.27Th_{sh} + 0.92Th_s + 4.19$	0.83	25.2	$p < 0.001$

eTh_x : Estimated hearing threshold for frequency x Hz.

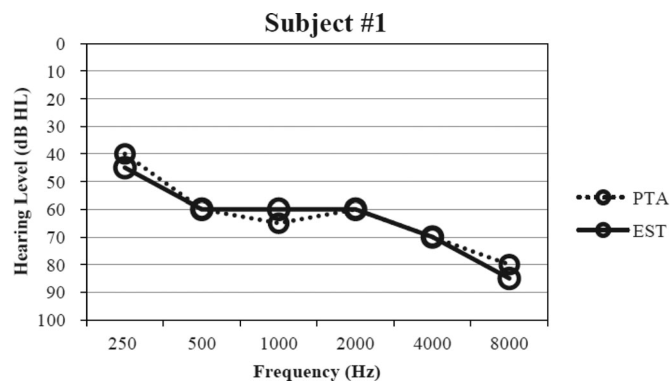

Fig. 2. Hearing thresholds estimated by conventional PTA and PhoSHA on a smartphone.

Since we did not have more measurements, we could not test these formulas with new data and analyze how well they fit. We instead chose 24 of the measurements to derive the coefficients and applied the resulting equation to predict the remaining measurement. We thus obtained 25 test results, which we discuss in the next section.

III. RESULTS

A. *Comparison Between PTA and PhoSHA*

As described in the previous section, 25 ears from 15 hearing-impaired subjects were tested at six frequencies. Therefore, there were 150 points of comparison between PTA and our proposed method. Fig. 2 shows the comparison for the right ear of subject #1. We observe a 5 dB HL difference at 250, 1000, and 8000 Hz, and no difference at 500, 2000, and 4000 Hz.

Fig. 3 shows the mean differences between the hearing thresholds estimated by PTA and those calculated using our method and the formulas in Table I. The mean difference is higher than 5 dB HL at low frequencies (250 and 500 Hz), but close to 5 dB HL at the other frequencies. The overall mean value across the six frequencies is 5.6 dB HL. The standard deviations for the six tested frequencies were 5.4, 6.7, 5.3, 4.1, 5.8, and 4.1 dB HL, respectively.

Of the 150 measurements, 29 measurements differed by 10 dB HL, 14 by 15 dB HL, and 4 by 20 dB HL. The maximum difference was 20 dB HL. The remaining 103 measurements differed by 5 dB HL or less (53 differed by 5 dB HL, and 50 were exactly correct).

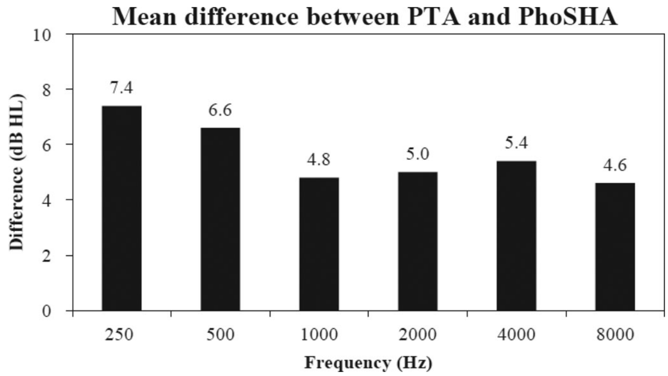

Fig. 3. Mean difference between hearing thresholds estimated by conventional PTA and PhoSHA.

TABLE II
MEAN TEST TIME

Group	Num. of Ears	Age in years Mean (Std. Dev.)	Mean Test Time (s)	Std. Dev. (s)
Hearing-impaired	25	69.9 (12.1)	95.4	44.1
Normal hearing	31	28.9 (5.0)	24.7	3.0

TABLE III
MEAN TEST TIME FOR EACH EAR

Group	Num. of Subjects	First Ear Test Time (s) Mean (Std. Dev.)	Second Ear Test Time (s) Mean (Std. Dev.)	Paired t Test
Hearing-impaired	20	101.2 (56.4)	91.1 (29.6)	p > 0.05
Normal hearing	14	26.2 (3.0)	23.8 (2.5)	p < 0.01

B. Test Time for PhoSHA

Table II shows the mean and standard deviation of the time taken by the hearing-impaired group and the normal hearing group to conduct a PhoSHA test. The mean test time for the hearing-impaired group was 95.4 s, but the standard deviation was 44.1 s. The mean test time for the normal hearing group was 28.9 s and the standard deviation was only 3.0 s.

Table III shows the mean test time of the first ear and the second ear to be tested. The first ear to be tested was the ear the subject typically uses for a phone call. The mean test time for the first ear of normal hearing subjects was 26.2 s, and that for their second ear was 23.8 s. There was a significant difference between the first-ear and second-ear times in the normal hearing subjects ($p < 0.01$). However, there was no significant mean test time difference between the first and second ear in the hearing-impaired group.

Fig. 4 shows the average test time for each test sound in the two ears. The test time for the second ear is shorter than that for the first ear. (23.8 and 26.1 s) Also, the test time for each phoneme is seen to decrease as the subjects progress through the phonemes because we adjusted the initial sound level for the phoneme as we mentioned in Section II.

C. Comparing Individual PTA and PhoSHA Measurements

The result in Fig. 3 was calculated using regression formulas derived using the whole dataset. As described in Section II, we need to verify the regression formulas using new data. Using 24 of the 25 measurements, we derived a set of regression coefficients and applied the resulting formula to predict the remaining measurement. This gave us 25 predictions for each frequency. By comparing the predictions with the results of PTA, we found that the mean differences were 9.4, 8.8, 7.0, 6.4, 6.0, and 6.4 dB HL at 250, 500, 1000, 2000, 4000, and 8000 Hz, respectively. All the differences were greater than 5 dB HL, and the overall mean difference across the six frequencies was 7.3 dB HL. The standard deviations were 6.8, 7.9, 6.8, 4.9, 6.3, and 4.9 dB HL, respectively.

Of the 150 predictions, 23 differed by 10 dB HL, 22 by 15 dB HL, 9 by 20 dB HL, and 3 by 25 dB HL. The maximum difference was 25 dB HL. The remaining 93 predictions differed by less than 5 dB (57 differed by 5 dB HL and 36 were exactly correct).

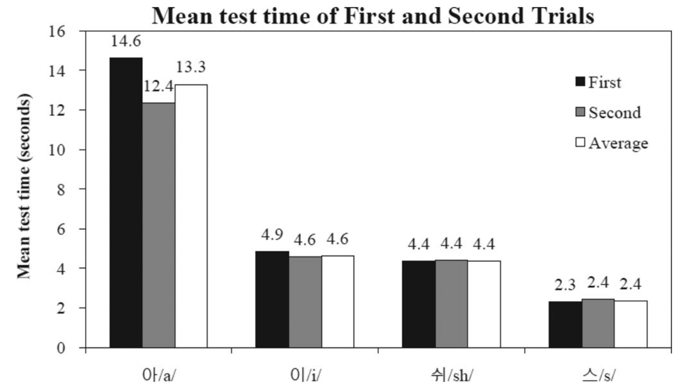

Fig. 4. Average test time of each phoneme in the first, second, and averaged ear test.

IV. DISCUSSION

Subjects under 50 years of age understood our experimental procedures and were able to perform the test themselves. On the other hand, most subjects over 50 years of age found it difficult to perform the test using a smartphone. As seen in Table II, there was a difference in age between normal hearing and hearing-impaired groups. The normal hearing subjects were generally younger than the hearing-impaired subjects. Therefore, we must consider the age difference when comparing mean test times between the two groups, although we concluded that there is a statistically significant difference in test time between the normal hearing group and the hearing-impaired group.

We chose four phonemes (two vowels, /a/ and /i/, and two consonants, /sh/ and /s/) considering their frequency spectra, but it was one of possible combinations. When we excluded one of two consonants in the calculation of regression coefficients, there was only about 10% accuracy reduction. It might be because that the frequency spectra of two consonants were overlapped. Two vowels - /a/ and /i/ showed very different frequencies of first formant and second formant, and it might be a positive effect to the resolution of regression analysis.

Table III shows the mean test time difference between the first ear and second ear, which may be caused by the learning effect. When the subject starts the test for the first time, his response to sound stimuli is relatively slow. After the subject becomes familiar with the system, his response becomes faster. We found a statistically significant mean test time difference between the first ear and the second ear. However, in hearing-impaired subjects, this effect was not critical because the hearing thresholds of their ears differed more than those of normal hearing subjects. Because the subject used the better ear as the first ear, the learning effect could not reduce the test time of the second ear. Another solution to eliminate the learning effect may be to play phonemes randomly.

PhoSHA was conducted in a sound-proof booth. In the real world, it may be conducted without a sound-proof booth in the presence of ambient noise. While this could pose a problem, the canal-type earphones with ear protection muff we use do reduce background noise to a certain extent.

Our method to determine hearing thresholds was a simplified version of the conventional rule of testing at least three times at a candidate level. This was one of main reasons for the dramatic reduction in mean test time for normal hearing subjects. If we implement the conventional rule such as two responses of three trials, the mean test time might increase to approximately twice its current duration.

It is very important to consider the properties of the earphone used. Different earphones are seen to give very different results due to their acoustic properties. Moreover, there was a difference between left and right earpieces. During calibration, we realized that four different Samsung smartphones (SHW-M250S) did not differ much at /a/ 70 dB HL, but four different earphones differed substantially. Although the performance of smartphone audio codec chips is rapidly improving, the problem of background noise becomes acute when we try to generate low-intensity sound stimuli below 25 dB HL. Therefore, the estimation of the hearing thresholds of normal hearing subjects whose thresholds were usually below 10 dB HL was difficult. For this reason, we performed the experiments of hearing thresholds with hearing impaired subjects. The thresholds of normal hearing subjects were all 30 dB HL and their test time data were only analyzed.

One set of regression formulas was calculated from all 25 hearing-impaired measurements. To verify how well these formulas work, it is advisable to test them on new data. Therefore, we chose 24 measurements as training data to calculate new sets of regression formulas and used these to predict the measurement for the one remaining ear. This is a method of cross validation which is generally used in the analysis of limited number of data. We obtained 25 predictions using this strategy and compared them with PTA results. The estimates by testing analysis were not as good as those obtained from the first set of regression formulas (training analysis), but the mean error was only 7.3 dB HL, as shown in Fig. 5. This value was higher than that in Fig. 3, which was 5.6 dB HL.

V. Conclusion

We proposed a new method of hearing assessment using phonemes and implemented it on a smartphone to allow hearing-impaired people to perform self-hearing tests. Regression formulas to estimate the PTA audiogram for six frequencies using PhoSHA with four Korean phonemes were successfully applied. When we did a cross-validation test, the difference was not much higher. In addition, the mean test time was reduced significantly because we used four phonemes instead of six pure tones. We conclude that PhoSHA is suitable for the estimation of audiograms with sufficient accuracy and reduced testing time.

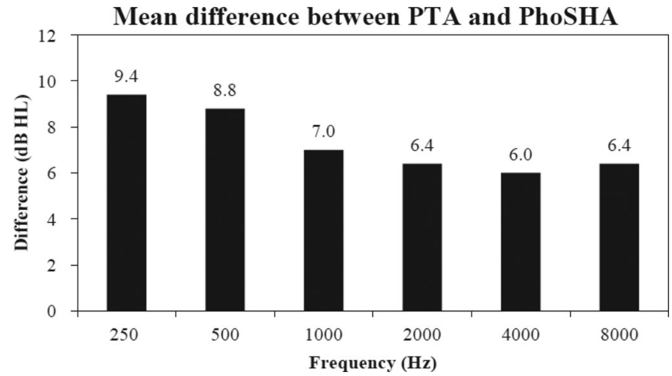

Fig. 5. Mean difference between hearing thresholds estimated by conventional PTA and PhoSHA with 25 regression formulas.

References

[1] M. N. Kamel Boulos, S. Wheeler, C. Tavares, and R. Jones, "How smartphones are changing the face of mobile and participatory healthcare: An overview, with example from eCAALYX," *Biomed. Eng. Online*, vol. 10, pp. 10–24, Apr. 2011.

[2] I. Bisio, F. Lavagetto, M. Marchese, and A. Sciarrone, "Smartphone-based user activity recognition method for health remote monitoring applications," in *Proc. 2nd Int. Conf. Pervasive Embedded Comput. Commun. Syst.*, Feb. 2012, pp. 200–205.

[3] H. Shim, J. H. Lee, S. O. Hwang, H. R. Yoon, and Y. R. Yoon, "Development of heart rate monitoring for mobile telemedicine using smartphone," in *Proc. 13th Int. Conf. Biomed. Eng.*, 2009, vol. 23, pp. 1116–1119.

[4] A. T. Ho, A. J. Hildreth, and L. Lindsey, "Computer-assisted audiometry versus manual audiometry," *Otol. Neurotol.*, vol. 30, pp. 876–883, Oct. 2009.

[5] W. Swanepoel de and L. Biagio, "Validity of diagnostic computer-based air and forehead bone conduction audiometry," *J. Occup. Environ. Hyg.*, vol. 8, pp. 210–214, Apr. 2011.

[6] J. M. Choi, H. B. Lee, C. S. Park, S. H. Oh, and K. S. Park, "PC-based tele-audiometry," *Telemed. J. E Health*, vol. 13, pp. 501–508, Oct. 2007.

[7] L. Honeth, C. Bexelius, M. Eriksson, S. Sandin, J. E. Litton, U. Rosenhall, O. Nyrén, and D. Bagger-Sjöbäck, "An internet-based hearing test for simple audiometry in nonclinical settings: Preliminary validation and proof of principle," *Otol. Neurotol.*, vol. 31, pp. 708–714, Jul. 2010.

[8] N. Nakamura, "Development of "MobileAudiometer" for screening using mobile phones," in *Proc. IEEE Eng. Med. Biol. Soc. Conf.*, 2004, vol. 5, pp. 3369–3372.

[9] B. A. Goldstein and C. W. Newman, "Clinical masking: A decision-making process," in *Handbook of Clinical Audiology*, 4th ed., J. Katz, Ed. Baltimore, MD, USA: Williams & Wilkins, 1994, pp. 109–131.

[10] R. J. Roeser, K. A. Buckley, and G. S. Stickney, "Pure tone test," in *Audiology Diagnosis*, 1st ed., R. J. Roeser, M. Valente, and H. Hosford-Dunn, Eds. New York, USA: Thieme, 2000, pp. 227–256.

[11] M. J. Middelweerd, J. M. Festen, and R. Plomp, "Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram," *Audiology*, vol. 29, no. 1, pp. 1–7, 1990.

[12] K. L. Payton, R. M. Uchanski, and L. D. Braida, "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," *J. Acoust. Soc. Am.*, vol. 95, no. 3, pp. 1581–1592, Mar. 1994.

[13] R. Carhart, "Observations on relations between thresholds for pure tones and for speech," *J. Speech Hear Disord.*, vol. 36, no. 4, pp. 476–483, Nov. 1971.

[14] J. Sohn, D. Kim, Y. Ku, K. Lee, and J. Lee, "Study on self hearing assessment using speech sounds," in *Proc. IEEE Eng. Med. Biol. Soc. Conf.*, 2011, pp. 2384–2387.

Authors' photographs and biographies not available at the time of publication.