Acoustic Features and Cortical Auditory Evoked Potentials according to Emotional Statues of /u/, /a/, /i/ Vowels

Chunhyeok Kim¹, Seungwan Lee¹, Inki Jin^{1,2}, and Jinsook Kim^{1,2}

¹Department of Speech Pathology and Audiology, Graduate School, Hallym University, Chuncheon,

²Division of Speech Pathology and Audiology, Research Institute of Audiology and Speech Pathology, College of Natural Sciences, Hallym University, Chuncheon, Korea

Received September 14, 2017

Revised October 19, 2017

Accepted November 6, 2017

Address for correspondence

Jinsook Kim, FAAA, PhD
Department of Speech Pathology and
Audiology, Graduate School,
Hallym University,
1 Hallymdaehak-gil,
Chuncheon 24252, Korea
Tel +82-33-248-2213
Fax +82-33-256-3420
E-mail jskim@hallym.ac.kr

Background and Objectives: Although Ling 6 sounds are often used in the rehabilitation process, its acoustic features have not been fully analyzed and represented in cortical responses. Current study was aimed to analyze acoustic features according to gender and emotional statuses of core vowels of Ling 6 sounds, /u/, /a/, and /i/. Cortical auditory evoked potentials (CAEPs) were also observed in those vowels. **Subjects and Methods:** Vowel sounds /u/, /a/, and /i/ out of Ling 6 sounds representing low, middle and high frequencies were recorded from normal 20 young adults. The participants watched relevant videos for 4–5 minutes in order for them to sympathize emotions with anger (A), happiness (H), and sadness (S) before producing vowels. And without any emotional salience, neutrally production was performed. The recording was extracted for 500 ms to select pure vowel portion of production. For analysis of CAEP, the latencies and amplitudes of P1, N1, P2, N2, N1-P2 were analyzed. **Results:** Intensities of /u/, /a/, and /i/ were 61.47, 63.38, and 60.55 dB. The intensities of neutral (N), H, A, S were 60.60, 65.43, 64.21, and 55.75 dB for vowel /u/, vowel /a/ were 61.80, 68.98, 66.50, and 56.23 dB, and vowel /i/ were 59.34, 64.90, 61.90, and 56.05 dB. The statistical significances for vowel and emotion were found but not for gender. The fundamental frequency (F0) of vowels for N, A, H, and S were 168.04, 174.93, 182.72, and 149.76 Hz and the first formant were 743.75, 815.59, 823.32, and 667.62 Hz. The statistical significance of F0 was found by vowel, emotion, and gender. The latencies and amplitudes of CAEP components did not show any statistical significance according to vowel. **Conclusions:** Ling 6 sounds should be produced consistently in the rehabilitation process for considering their difference of intensities and frequencies according to speaker's emotions and gender. The vowels seemed to be interpreted as tonal stimuli for CAEP components of this study with similar acoustic features among them. Careful selection of materials is necessary to observe meaningful conclusion of CAEP measurement with vowel stimuli.

J Audiol Otol 2018;22(2):80-88

KEY WORDS: Acoustic feature · Vowel · Emotion · Cortical auditory evoked potential.

Introduction

Speech is a complex sound and can be analyzed by the intensity and frequency. These acoustic features of speech sound depend on gender, phoneme, content of conversation, and

emotion. It is reported that males generally conduct much lower sound than females do and adults conduct much lower sound than youth do [1]. The average fundamental frequency (F0) of the articulation of adults is 222.9 Hz [2], with the averages of F0s of female and male are 216.7 Hz and 112.1 Hz, showing 1.8 times higher frequency in female [3]. In a reading distance of 15 cm from a microphone, the mean intensity of person's speech was measured at 69.3 dB SPL [4], and the mean intensity of conversation was measured at 50 to 70 dB

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

HL in general [5]. To be heard comfortably across speaking environments, it should be ranged from 63 to 80 dB among normal adults [6].

Acoustic features of phonemes are also dependent on gender and emotional statuses. For example, F0s of /a/, /i/, and /o/ were 127, 143, and 129 Hz in male and 214, 230, and 215 Hz in female, respectively [7]. Other than F0, the characteristic of frequency of vowel can also be described by the vowel chart with the vowel trapezium. Three categories, high, middle, low vowels are classified by the tongue position when a speaker articulates corresponding vowels. High vowels are /i/, /u/, /e/, and /ʊ/, middle vowels are /o/, /ɪ/, and /æ/, and low vowels are /a/, /ɛ/, /ʌ/, and /æ/, with International Phonetic Alphabet symbols. The intensities of /a/, /i/, and /o/ were 78.5, 77.5, and 80.2 dB in male, 73.2, 72.3, and 75 dB in female, at a recording distance of 10 cm from a microphone [7].

The emotional status is one of the factors that can change acoustic features because of prosodic features of the different emotions. Yildirim, et al. [8] investigated acoustic properties of sentences uttered by a semi-professional actress associated with four different emotions such as sadness (S), anger (A), happiness (H), and neutral (N). The means of F0s were 188, 195, 233, and 237 Hz for N, S, A, and H. Authors suggested that the positioning of tongue for speech production can possibly differentiate frequencies for the various emotional statuses. They concluded that happy/anger and neutral/sad share similar acoustic features. Higher pitch and wider energy values were found in happy and anger emotions. This phenomenon was exhibited not only in speech samples but also in phonemes [9].

The Ling 6 sounds were composed of /u/, /a/, /i/, /sh/, /s/, and /m/ [10]. As six sounds have concentrated energies across the speech frequency range from 250 Hz to 8,000 Hz, identification of these sounds is known to be meaningful in estimating hearing thresholds at relevant frequencies for children. In Korean, the average of /u/ was measured at 300 Hz with the range from 284 Hz to 519 Hz, /a/ at 1,500 Hz with the range from 659 Hz to 2,314 Hz and /i/ at 3,200 Hz with the range from 2,046 Hz to 3,367 Hz, and /sh/ and /s/ were measured at about 4,000 Hz to 6,000 Hz [11]. Although the frequency range of those words covered from 250 Hz to 8,000 Hz which was thought to be the usual speech frequency range [12], the frequencies of individual words, specially the vowels showed different frequency values across the languages [13-15]. And if the emotional and gender factors are added, the variability could be more complicated. Nevertheless, these sounds are widely used as a fast and easy assessment tool for both hearing check and estimation of speech perception area for the aural rehabilitation process for children, the acoustic

features of these sounds are of interest among researchers.

The cortical auditory evoked potentials (CAEPs) including P1, N1, P2, and N2 is also called the long latency auditory evoked potentials, as it typically measures the electrical activity that occurs in the central auditory system at about 50 ms after stimulation. The normal presence of CAEP components indicates that the signal was encoded in the auditory cortex properly. Also, these can reflect the neural activity of the brain at the different levels. Out of the components of CAEP, P1 represents the activity of secondary auditory cortex and Heschl's gyrus in the auditory cortex and N1 represents the activity of primary auditory cortex by the activities of left and right hemispheres [16]. Therefore many investigations were performed to verify the origins presuming higher central nervous system and to compare the responses of the normal hearing to the patients with autism, attention deficit disorder, and hearing impairment [17,18]. Different acoustic features are valuable factors for analyzing CAEPs showing larger amplitudes when using speech stimuli than when using tone bursts implying superior cortical processing of the speech stimuli [19]. Further, components of CAEP were generally affected by the intensity and frequency of the stimuli. As the intensity and frequency increases, the amplitude increases and the latency decreases with tonal stimuli [20]. With speech stimuli, the intensity effect remained identical [21,22], while the frequency effect was scarcely known since speech stimuli were composed of multi frequency components. Interestingly, it was reported that N1 latency decreased as F0 increased for complex stimuli when the frequency was more than 62.5 Hz apart [23]. A previous research reported N100m which is thought to be N1 component in CAEP can be changed by different vowels such as /a/, /i/, and /e/, because of different acoustic features [24]. Conclusively, CAEPs can be utilized as objective indicators of changes to acoustical features of speech including vowels. And these results can let us have more comprehensive knowledge about central auditory nervous system with decoding speech.

The purpose of this study was to evaluate and analyze the acoustic features according to four emotions, N, A, H, and S of three vowels, /u/, /a/, and /i/ out of Ling 6 sounds. These vowels were selected as /u/, /a/, and /i/ representing low, middle and high frequencies of hearing and core vowels for rehabilitation. Especially, the frequency of those vowels were different across the languages, identifying Korean frequency ranges of those vowels would be meaningful for better use of Ling 6 sound. We hypothesized that frequencies and intensities out of several acoustic features be changed by gender, emotional statuses and vowels themselves. Additionally, as CAEPs be different according to different vowels for reflecting the acous-

tic features of /u/, /a/, and /i/, the CAEPs were explored for identifying as possible objective windows of the corresponding vowels.

Subjects and Methods

A total of 20 young adults, 10 males and 10 females (mean: 22.1, standard deviation: 2.17) produced /u/, /a/, and /i/ vowels according to emotions of A, H, S, and N. The subjects watched relative emotional screens for 4 to 5 minutes prior to producing vowels so that they can draw their empathy for production of vowels for the corresponding emotions. The videos were selected in the Youtube. N was recorded without watching the video. An angry emotion was induced right after watching the video about animal abuse (<https://www.youtube.com/watch?v=olDsYSou7fQ>), a happy emotion was induced right after watching the video about fun tumbling mistakes which make people laugh (<https://www.youtube.com/watch?v=u3XWq63IMFE>), and a sad emotion was induced right after watching the video about regretting for not being nice for parents (<https://www.youtube.com/watch?v=p-qZ76akjKE>). Considering, the age of subjects, category, and checking numbers, the specific videos were chosen for inducing appropriate emotions when producing vowels from the group of investigators.

The participants produced a sentence including the target vowel for consistent analysis of vowel sound part. For example, "I speak /a/" sentence was recorded for the target vowel /a/. The production of vowel part of sentence was sustained for 2 seconds. Only well-recorded portion was extracted for 500ms in order to control the quality of the data. When each recording session was finished, a 3 minutes break was given to the participant for separating the emotional status from the last record. The device used for recording and analysis was the main program of Computerized Speech Lab (CSL) (KayPENTAX™, Lincoln Park, NJ, USA). Also, Multi-Dimensional Voice Program (KayPENTAX™, Lincoln Park, NJ, USA) was used to identify the normal range of hoarseness and vibration of the production. Also, the confusion matrix was evaluated to find out how effectively the acoustic features discriminate emotions. 10 young adults, 5 males and 5 females (mean: 23.8, standard deviation: 2.11), participated for evaluation of confusion matrix. Twenty sets of stimuli composed with 5 of 4 emotional sorts were evaluated for each vowel. The participants selected the emotions that they think they heard. Then their selections were matched for the correct emotions provided.

Pure tone audiometry was performed to all participants for confirming normal hearing prior to the CAEP measurements. The CAEP responses were measured from 18 normal young

adults, 9 males and 9 females, with age average of 22.3 (male: 22.6, female: 22.0). Bio-logic Navigator Pro System (Natus, Mundelin, IL, USA) was utilized as an experimental device. Out of recorded production of vowel sounds, neutral recordings of /u/, /a/, and /i/ vowels were selected for this experiment. The closest mean frequency and intensity productions of male and female were used. Male participants listened recorded vowel sounds of male and female participants listened recorded vowel sounds of female as stimuli. Acoustic features of vowel used to CAEP for male and female were shown in Table 1. The stimulus was presented at 70 dB HL using rarefaction polarity with the rate of 0.7 stimuli per second. 150 responses were taken for each recorded vowel. Electrodes were attached in two channels with gold cup electrodes, non-inverting electrode was attached to Fz, common to Fpz, and inverting electrode, to both right and left mastoids. The band-pass filter was from 1 to 30 Hz and gain was 10,000 for both channels. CAEP's components, P1, N1, P2, N2, and amplitude of N1-P2 complex were measured.

For analyzing acoustic features of vowels statistically, the independent variables were 3 vowels, /u/, /a/, and /i/, 4 emotions, N, A, H, and S, and gender, male and female. The dependent variables were intensity and frequency of the vowels. Three ways mixed analysis of variance (ANOVA) was utilized for the statistical analysis. For the CAEP experiment, independent variable was 3 vowels, /u/, /a/, and /i/ and the dependent variables were latencies and amplitudes of the components of CAEP. One way ANOVA was utilized for the statistical analysis. The statistical package for the social sciences version 20.0 (IBM Corp., Armonk, NY, USA) was applied and the statistical analysis conducted at significant level of 0.05. The study was approved by the Institutional Review Board of Hallym University (HIRB-2015-090). The informed consent form was reviewed and approved by the IRB.

Table 1. Acoustic features of vowel stimuli for cortical auditory evoked potential measurements

Stimuli	Acoustic features			
	Intensity (dB)	F0 (Hz)	F1 (Hz)	F2 (Hz)
Male voice				
/a/	58.39	111.44	1,037.37	2,713.80
/u/	58.67	124.22	543.31	3,690.43
/i/	56.70	125.47	2,489.20	3,661.72
Female voice				
/a/	54.70	199.16	1,284.56	3,624.67
/u/	57.03	214.54	481.45	4,158.43
/i/	51.30	225.98	364.04	2,558.20

F0: fundamental frequency, F1: first formant, F2: second formant

Results

Acoustic features

Intensities were 62.29 dB and 61.31 dB for the male and female voices. Although the production of male participants showed stronger intensity than that of female participants with the difference of 0.982, there was no statistical significance ($p>0.05$). Additionally, spectrograms showed wider intensity range for the female's voice (Fig. 1).

Intensities of /u/, /a/, and /i/ were 61.47 dB, 63.38 dB, and 60.55 dB, revealing the highest intensity at /a/ and the lowest intensity at /i/ with the statistical significance ($p<0.05$) (Table 1, 2), respectively. A Bonferroni post hoc test indicated that

Fig. 1. Spectrogram of vowels with various emotional statuses for 500 ms duration. The upper features represent male productions and the lower features represent female productions.

intensity of /a/ was significantly different from /i/ ($p<0.05$). The intensities by emotions of N, A, H, and S were 60.54 dB, 66.44 dB, 64.21 dB, and 56.01 dB revealing the highest intensity with A and the lowest intensity with S showing the statistical significance ($p<0.05$). The strongest intensity was recorded with A followed by H, N, and S in the order. Interaction effect showed the statistical significance between vowel and emotion ($p<0.05$) (Fig. 2). The highest intensity was recorded by /a/ with A in the female's production.

Frequencies of the produced vowels were analyzed by the F0, the first formant (F1), the second formant (F2), and the third formant (F3) sequentially (Fig. 3). F0s were 139.12 Hz and 198.60 Hz for the male and female voices with the statistical significance ($p<0.05$). F0s of /u/, /a/, and /i/ were 171.69 Hz, 161.35 Hz, and 173.54 Hz with the statistical significance ($p<0.05$). A Bonferroni post hoc test indicated that F0s of vowels were significantly different in the pairs of /a/-/i/ and /a/-/u/ ($p<0.05$). F0s of N, A, H, and S were 168.04 Hz, 174.93 Hz, 182.72 Hz, and 149.76 Hz with the statistical significance ($p<0.05$). A Bonferroni post hoc test indicated that F0s of vowels were significantly different between S and the rest of emotions ($p<0.05$) (Table 3).

F1s were 730.45 Hz and 793.69 Hz for the male and female voices, respectively, with no statistical significance ($p>0.05$).

Fig. 2. Mean intensities of vowels according to the emotional statuses.

Table 2. Summary of separate three-factor mixed analysis of variance comparing differences of the produced intensity as a function of vowel, emotion, and gender

Factor	Sum of square	df	Mean square	F	p	η^2
Vowel	332.389	2	166.195	9.282	0.001*	0.340
Emotion	3,746.324	3	1,248.775	140.873	0.000*	0.887
Gender	57.859	1	57.859	0.478	0.498	0.026
Vowel×emotion	137.592	6	22.932	3.191	0.006*	0.151
Emotion×gender	13.621	3	4.540	0.512	0.676	0.028
Vowel×gender	101.856	2	50.928	2.844	0.071	0.136
Vowel×emotion×gender	50.975	6	8.496	1.182	0.321	0.062

* $p<0.05$

Fig. 3. F0, F1, F2, and F3 for each vowel according to different emotions. F0: fundamental frequency, F1: first formant, F2: second formant, F3: third formant.

Table 3. Summary of statistical analyses comparing differences of mean F0 and the F1 frequencies as a function of vowel, emotion, and gender

Factor	Sum of square	df	Mean square	F	p	η^2
F0						
Vowel	6,905.616	2	3,452.808	6.665	0.003*	0.270
Emotion	35,670.341	1.822	19,582.848	10.295	0.000*	0.364
Gender	212,230.590	1	212,230.590	94.536	0.000*	0.840
Vowel $\times$ emotion	1,598.830	3.219	496.742	0.646	0.599	0.035
Emotion $\times$ gender	77,918.202	1.822	42,776.723	22.489	0.000*	0.555
Vowel $\times$ gender	2,206.616	2	1,103.308	2.130	0.134	0.106
Vowel $\times$ emotion $\times$ gender	6,471.248	3.219	2,010.556	2.614	0.056	0.127
F1						
Vowel	16,826,171.805	2	8,413,085.902	39.069	0.000*	0.685
Emotion	956,982.840	3	318,994.280	2.501	0.069	0.122
Gender	239,959.753	1	239,959.753	1.859	0.190	0.094
Vowel $\times$ emotion	953,178.986	6	158,863.164	1.353	0.240	0.070
Emotion $\times$ gender	238,118.667	3	79,372.889	0.622	0.604	0.033
Vowel $\times$ gender	205,478.792	2	102,739.396	0.477	0.624	0.026
Vowel $\times$ emotion $\times$ gender	888,911.775	6	148,151.962	1.262	0.281	0.066

* $p < 0.05$. F0: fundamental frequency, F1: first formant

F1s of /u/, /a/, and /i/ were 740.55 Hz, 1,096.58 Hz, and 449.07 Hz with the statistical significance ($p < 0.05$). A Bonferroni post hoc test indicated that F1s of vowels were significantly different for pairs of /a/–/i/, /a/–/u/, and /i/–/u/ ($p < 0.05$). F2s were 3,146.93 Hz and 3,241.85 Hz for the male and female voices with no statistical significance ($p > 0.05$). F2s of /u/, /a/, and /i/ were 3,761.80 Hz, 2,975.15 Hz, and 2,846.20 Hz with no statistical significance ($p > 0.05$).

Table 4. Confusion matrix for discriminant analysis of the emotional salience

Emotion	Neutral	Anger	Happiness	Sadness	Other
/u/ (%)					
Neutral	46 (92)	-	-	4 (8)	-
Anger	7 (14)	34 (68)	1 (2)	8 (16)	-
Happiness	1 (2)	1 (2)	45 (90)	2 (4)	1 (2)
Sadness	-	-	-	50 (100)	-
/a/ (%)					
Neutral	48 (96)	-	-	2 (4)	-
Anger	5 (10)	40 (80)	3 (6)	2 (4)	-
Happiness	3 (6)	-	45 (90)	2 (4)	-
Sadness	3 (6)	-	1 (2)	44 (88)	2 (4)
/i/ (%)					
Neutral	49 (98)	-	-	1 (2)	-
Anger	-	37 (74)	13 (26)	-	-
Happiness	-	1 (2)	49 (98)	-	-
Sadness	6 (12)	-	3 (6)	41 (82)	-

F2s were 3,345.52 Hz, 2,987.79 Hz, 3,014.26 Hz, and 3,429.97 Hz for N, A, H, and S with no statistical significance ($p > 0.05$). F3s were 5,065.51 Hz and 4,906.35 Hz for the male and female voices with no statistical significance ($p > 0.05$). F3s of /u/, /a/, and /i/ were 5,858.29 Hz, 5,051.93 Hz, and 4,047.57 Hz with the statistical significance ($p < 0.05$). The Bonferroni with post hoc test indicated that the statistical significance between /i/ and the others. F3s were 5,083.74 Hz, 4,519.23 Hz, 4,789.74 Hz, and 5,550.99 Hz for N, A, H, and S with the statistical significance ($p < 0.05$) (Fig. 3).

The confusion matrix analysis was applied to evaluate the effectiveness of discriminating emotions (Table 4). The matrices indicated that the acoustic features were relevant to distinguish the corresponding emotion revealing 68–100% correction probability. Angry emotion expressed seemed harder

Fig. 4. Cortical auditory evoked potentials waveforms according to vowels, /u/, /a/, and /i/.

Table 5. Latencies and amplitudes of the Cortical auditory evoked potentials as function of gender and vowel

Vowel	P1		N1		P2		N2		N1-P2
	Latency (ms)	Amplitude (μV)	Latency (ms)	Amplitude (μV)	Latency (ms)	Amplitude (μV)	Latency (ms)	Amplitude (μV)	Amplitude (μV)
Male									
/u/	41.54	1.53	102.38	-3.60	193.99	2.13	252.28	-1.03	5.74
/a/	35.75	1.77	93.82	-3.83	172.25	3.03	255.87	-1.24	7.08
/i/	48.71	1.39	103.19	-3.84	179.64	2.46	266.51	-1.56	6.31
Total	42.00	1.56	99.79	-3.75	181.96	2.54	258.22	-1.27	6.37
p-value	0.18	0.47	0.05	0.78	0.39	0.22	0.54	0.30	0.20
Female									
/u/	45.81	1.22	102.71	-2.35	160.67	1.50	241.74	-1.53	4.00
/a/	45.93	1.04	100.06	-2.72	153.39	2.02	238.93	-1.83	4.74
/i/	53.10	1.27	106.66	-3.05	170.27	1.84	239.44	-1.44	4.90
Total	48.28	1.17	103.14	-2.70	161.44	1.78	240.03	-1.60	4.54
p-value	0.47	0.66	0.51	0.43	0.16	0.44	0.98	0.68	0.23
Total									
/u/	43.67	1.37	102.54	-2.97	177.32	1.81	247.01	-1.30	4.86
/a/	40.84	1.40	96.93	-3.27	162.81	2.52	247.39	-1.53	5.90
/i/	50.90	1.33	104.92	-3.44	174.95	2.14	252.97	-1.49	5.60
Total	45.14	1.37	101.47	-3.23	171.69	2.16	249.12	-1.44	5.57
p-value	0.10	0.94	0.06	0.35	0.23	0.09	0.79	0.67	0.07

to distinguish showing the lowest correction rate among emotions showing 68–80% correction rate while others showing 82–100% correction rate.

CAEPs

Each vowel revealed a different waveform in CAEP recordings. Fig. 4 shows average waveforms of CAEPs by vowels, /u/, /a/, and /i/.

P1 latencies and amplitudes of /u/, /a/, and /i/ were 43.67, 40.84 and 50.90 ms and 1.37, 1.40, and 1.33 μ V. N1 latencies and amplitudes of /u/, /a/, and /i/ were 102.54, 96.93, and 104.92 ms and -2.97, -3.27, and -3.44 μ V. P2 latencies and amplitudes of /u/, /a/, and /i/ were 177.32, 162.81, and 174.95 ms and 1.81, 2.52, and 2.14 μ V. N2 latencies and amplitudes of /u/, /a/, and /i/ were 247.01, 247.39, and 252.97 ms and -1.30, -1.53, and -1.49 μ V. The N1-P2 amplitude of /u/, /a/, and /i/ were 4.86, 5.90, and 5.60 μ V (Table 5). Any of CAEP components did not show statistical significances for latencies and amplitudes according to vowels for both male and female. However general tendency could be observed by the intensity of stimulus. As the intensities increased (/i/ < /a/ < /u/), the latencies decreased and the amplitude increased. Any of the frequency characteristics did not affect to the latency and amplitude of CAEP components of this study.

Discussion

When vowel sounds of /u/, /a/, and /i/ representing low, middle, and high frequencies out of Ling 6 sounds were analyzed by CSL. The average intensity of the total was 61.80 dB showing similar result of 58.30 dB to the previous study [25]. Intensities for the male and female voices were 62.29 dB and 61.31 dB in the present study, showing similar results with previous studies [6,7]. In the present study, the intensity of /a/ was the strongest showing consistent results to the preceding research [7]. The order of intensity of this study was /a/, /u/, and /i/ while Brockmann, et al. [7] reported order of intensity as /o/, /a/, and /i/. As the present study did not include vowel /o/ and the other study did not include vowel /u/, we can conclude /a/ showed stronger intensity than /i/ consistently.

ly. When the intensities by emotions were compared, the intensity of A and H showed stronger than N and S for the present study and the preceding study [8].

The average of the F0s of the vowels of the present study was 168.86 Hz and the average of F0s of the sentence was 178.13 Hz in the previous investigation [26]. Both showed similar results. F0s for the male and female voices of this study were 139.12 Hz and 198.60 Hz and the preceding study showed the range 92.8–156.1 Hz for the male voices and 172.4–274.5 Hz for the female voices. All recorded about 79.6–118.4 Hz higher frequencies for the female voices. This phenomenon has been widely agreed among investigations, as many researches also reported that the vowel productions of the male were lower and stronger than those of the female [4,7,27]. The stronger intensity with lower frequency of the male voice can be interpreted by the size of the male's vocal fold as many investigators explained [27,28].

The frequency of vowel in terms of F0 seemed to be affected by tongue position in oral cavity when the vowel was produced because /u/ and /i/ showed higher frequencies consistently across the investigations [28,29]. The fact that /u/ and /i/ are sorted as high vowels at the vowel chart which is classified by the characteristic of tongue position supports this issue. For example, F0s of /u/ and /i/ were 171.69 Hz and 173.54 Hz showing higher than 161.35 Hz of /a/ in the present study. Interestingly, the order of pitches of the vowel, /u/, /i/, and /a/ was found to be similar across the languages, although the frequencies of the vowels were affected by the linguistic characteristics. Whalen and Levitt [29] compared 32 different languages F0s of /u/, /a/, and /i/ and reported that they all showed different values. However, the order of pitches in frequency with /u/, /i/, and /a/ remained the same. Additionally, F1s and F2s of /u/, /a/, and /i/ seemed to be affected by different languages. When, French, German, Australian English and Korean with the results of the present study were compared for the female voices of F1, the largest difference was revealed in /a/ and the smallest difference was revealed in /i/ across the languages. But the higher frequencies in of /u/ and /a/ were also observed (Table 6) [13–15].

The highest F0 was revealed with H and followed by A, N, and S in the order. And these results were similar to the previous studies [8], although the researcher used different materials from this study, for example, short sentences and two-syllables. The order of emotions of F0s was not affected by different materials. Taken all together, both H and A seemed to show the highest intensity and frequency regardless of materials utilized. Possibly, the reason could be the expressive way of extreme emotions which had strong acoustic qualities. Because A and H situated opposite to each other in the pleasure axis,

Table 6. The F1's and F2's of /u/, /a/, and /i/ in different languages

Language	/u/		/a/		/i/	
	F1	F2	F1	F2	F1	F2
French	404	1,105	685	1,677	348	2,365
German	350	1,048	779	1,347	329	2,316
Australian English	391	1,342	860	1,423	440	2,849
Present study	839	4,169	1,293	3,366	332	3,176

F1: first formant, F2: second formant

acoustic features were commonly observed strongly in other investigations [8,9].

The emotional matrix revealed that the average of correct rate was 87.9% of this study. This rate was higher than the value of 68.2% reported by the previous study [8]. This difference was thought to be due to the materials of the experiment between the vowels of this study and the sentences of the previous study.

Conclusively, subtle changes of acoustic features such as the intensity and frequency according to vowels and emotions should be considered when performing the Ling 6 sound test, although those sounds are within speech frequencies, 250–8,000 Hz. The speaker should be careful with the emotions when fitting the hearing aids and performing aural rehabilitations for children with vowels /u/, /a/, and /i/. Also, the gender of the speaker should also be considered. With the importance of the frequency range heard by children in the rehabilitation session, the consistency of the voice for Ling 6 sound test should be self-checked during the session. However, further investigation is recommended to evaluate the /m/, /s/ and /ʃ/ sounds in order to analyze the acoustic features of the whole Ling 6 sounds with emotions and gender.

For CAEPs, the vowels of this study did not show any statistical significance for the latency and amplitude for both male and female participants. However, general tendency of intensity was observed by the vowels, /u/, /a/, and /i/. As the intensity increased (/i/ > /a/ > /u/), the latencies decreased and the amplitude increased. The significant intensity effect was observed at the previous researches with more than 20 dB difference [21,22]. But the intensity difference was only 1.87 dB among the male's vowel stimuli and 5.73 dB among the female's vowel stimuli of this study. This would be the reason why there was no statistical significance depending on the intensity of the vowel stimuli for the present study. Also, N1 component which showed statistical difference in the frequency for the previous study [23] did not show any statistical significance in the present study. This discrepancy can also be explained by difference of the frequency utilized as materials. Because the research materials of the previous study had 62.5 Hz apart in frequency difference, while the materials of this study had less than 30 Hz apart among F0s vowels. For the reason, the frequency did not vary the latencies and amplitudes of CAEP components. The vowels used for this study seemed to act as tonal stimuli with tiny intensity and frequency differences. Therefore, the acoustic features of the vowels did not vary the latency and amplitude of CAEP components in the present study. However, some investigations observed that the spectral features of vowels reflected different source of locations of cortical responses meaningfully, when multichan-

nel parameters or fMRI were utilized [30,31]. It was implied that the categorized cortical area represented activated locations in auditory cortex depending on different vowel stimulations. Further study should be performed with multichannel parameters for measurements and careful selection of materials including dramatic difference of the intensity and frequency for recording reflections of acoustic features of various vowels. The present study did not include the emotional factors of the vowel stimuli for CAEP measurements which may lead us for better understanding of the exogenous and endogenous components of the CAEP results. This is one of the limitations of the present study and the following experiment is being performed reflecting emotional salience of the vowel stimuli for measuring CAEP.

Acknowledgments

This research was supported by Hallym University Research Fund, HRF-201410-014.

Conflicts of interest

The authors have no financial conflicts of interest.

REFERENCES

- 1) Aronson AE, Bless DM. Clinical voice disorders. 4th ed. New York, NY: Thieme Medical Publishers;2009. p.10-7.
- 2) Huh MJ, Jeong OR. Acoustic characteristics of prelingual hearing impaired speakers. *J Speech Hear Disord* 1997;6:61-77.
- 3) Nam DH, Rheem S, Choi HS. Correlation between the external laryngeal length and the habitual speaking fundamental frequency. *Phon Speech Sci* 2009;1:187-93.
- 4) Baken RJ, Orlikoff RF. Clinical measurement of speech and voice. 2nd ed. San Diego, CA: Singular Publishing Group;2000. p.108-9.
- 5) Hacki T. Comparative speaking, shouting and singing voice range profile measurement: physiological and pathological aspects. *Logoped Phoniatr Vocol* 1996;21:123-9.
- 6) Brockmann M, Storck C, Carding PN, Drinnan MJ. Voice loudness and gender effects on jitter and shimmer in healthy adults. *J Speech Lang Hear Res* 2008;51:1152-60.
- 7) Brockmann M, Drinnan MJ, Storck C, Carding PN. Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task. *J Voice* 2011;25:44-53.
- 8) Yildirim S, Bulut M, Lee CM, Kazemzadeh A, Busso C, Deng Z, et al. An acoustic study of emotions expressed in speech. *Proceeding of the 8th International Conference on Spoken Language Processing*; 2004 Oct 4-8: Jeju. Jeju: International Speech Communication Association;2004. p.2193-6.
- 9) Bachorowski JA, Owren MJ. Vocal expression of emotion: acoustic properties of speech are associated with emotional intensity and context. *Psychol Sci* 1995;6:219-24.
- 10) Ling D. The ling six-sound test. *Proceeding of the 2002 Alexander Graham Bell Convention*; 2002 Jun 29-Jul 2:St Louis, MO.
- 11) Kim JS, Lee KD, Ji YS. A study of the frequency analysis of the Korean meaningful monosyllabic words. *Audiol Speech Res* 2010;6:37-49.
- 12) Park H, Kim J. Comprehension and application of the Ling 6 sound test. *Audiol Speech Res* 2016;12:195-203.
- 13) Gendrot C, Adda-Decker M. Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news

- corpora in French and German. *Interspeech* 2005;2453-6.
- 14) Pätzold M, Simpson AP. Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. In: Simpson AP, Kohler KJ, Rettstadt T, editors. *The Kiel Corpus of Read/Spontaneous Speech-Acoustic data base, processing tools and analysis results*. Kiel: Universität Kiel; 1997. p.215-47.
- 15) Agung K, Purdy SC, McMahon CM, Newall P. The use of cortical auditory evoked potentials to evaluate neural encoding of speech sounds in adults. *J Am Acad Audiol* 2006;17:559-72.
- 16) Mäkelä JP, McEvoy L. Auditory evoked fields to illusory sound source movements. *Exp Brain Res* 1996;110:446-54.
- 17) Ceponiene R, Shestakova A, Balan P, Alku P, Ylaguchi K, Näätänen R. Children's auditory event-related potentials index sound complexity and "speechness". *Int J Neurosci* 2001;109:245-60.
- 18) Ceponiene R, Cummings A, Wulfeck B, Ballantyne A, Townsend J. Spectral vs. temporal auditory processing in specific language impairment: a developmental ERP study. *Brain Lang* 2009;110:107-20.
- 19) Van Dun B, Dillon H, Seeto M. Estimating hearing thresholds in hearing-impaired adults through objective detection of cortical auditory evoked potentials. *J Am Acad Audiol* 2015;26:370-83.
- 20) Hall JW. *New handbook of auditory evoked responses*. 1st ed. Boston, MA: Pearson;2007. p.490-2.
- 21) Prakash H, Abraham A, Rajashekar B, Yerraguntla K. The effect of intensity on the speech evoked auditory late latency response in normal hearing individuals. *J Int Adv Otol* 2016;12:67-71.
- 22) Purdy SC, Sharma M, Munro KJ, Morgan CL. Stimulus level effects on speech-evoked obligatory cortical auditory evoked potentials in infants with normal hearing. *Clin Neurophysiol* 2013;124:474-80.
- 23) Crottaz-Herbette S, Ragot R. Perception of complex sounds: N1 latency codes pitch and topography codes spectra. *Clin Neurophysiol* 2000;111:1759-66.
- 24) Obleser J, Eulitz C, Lahiri A, Elbert T. Gender differences in functional hemispheric asymmetry during processing of vowels as reflected by the human brain magnetic response. *Neurosci Lett* 2001;314:131-4.
- 25) Zraick RI, Marshall W, Smith-Olinde L, Montague JC. The effect of task on determination of habitual loudness. *J Voice* 2004;18:176-82.
- 26) Lee M. Variance characteristics of speaking fundamental frequency and vocal intensity depending on utterance conditions. *Phon Speech Sci* 2012;4:111-8.
- 27) Hwa Chen S. Sex differences in frequency and intensity in reading and voice range profiles for Taiwanese adult speakers. *Folia Phoniatr Logop* 2007;59:1-9.
- 28) Yang BA. comparative study of American English and Korean vowels produced by male and female speakers. *J Phon* 1996;24:245-61.
- 29) Whalen DH, Levitt AG. The universality of intrinsic F0 of vowels. *J Phon* 1995;23:349-66.
- 30) Obleser J, Lahiri A, Eulitz C. Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J Cogn Neurosci* 2004;16:31-9.
- 31) Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roetinger M, et al. Vowel sound extraction in anterior superior temporal cortex. *Hum Brain Mapp* 2006;27:562-71.