ORIGINAL ARTICLE
J Audiol Otol 2022;26(2):68-75

pISSN 2384-1621 / eISSN 2384-1710
https://doi.org/10.7874/jao.2021.00479

Effects of Natural Versus Synthetic Consonant and Vowel Stimuli on Cortical Auditory-Evoked Potential

Hyunwook Song[1], Seungik Jeon[1], Yerim Shin[1], Woojae Han[1,2], Saea Kim[1], Chanbeom Kwak[1], Eunsung Lee[1], and Jinsook Kim[1,2]

[1]Department of Speech Pathology and Audiology, Graduate School, Hallym University, Chuncheon, Korea
[2]Division of Speech Pathology and Audiology, College of Natural Sciences, Hallym University, Chuncheon, Korea

Received August 23, 2021
Revised October 21, 2021
Accepted November 9, 2021

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: Natural and synthetic speech signals effectively stimulate cortical auditory evoked potential (CAEP). This study aimed to select the speech materials for CAEP and identify CAEP waveforms according to gender of speaker (GS) and gender of listener (GL). **Subjects and Methods**: Two experiments including a comparison of natural and synthetic stimuli and CAEP measurement were performed of 21 young announcers and 40 young adults. Plosive /g/ and /b/ and aspirated plosive /k/ and /p/ were combined to /a/. Six bisyllables–/ga/-/ka/, /ga/-/ba/, /ga/-/pa/, /ka/-/ba/, /ka/-/pa/, and /ba/-/pa/–were formulated as tentative forwarding and backwarding orders. In the natural and synthetic stimulation mode (SM) according to GS, /ka/ and /pa/ were selected through the first experiment used for CAEP measurement. **Results**: The correction rate differences were largest (74%) at /ka/-/pa/ and /pa/-/ka/; thus, they were selected as stimulation materals for CAEP measurement. The SM showed shorter latency with P2 and N1-P2 with natural stimulation and N2 with synthetic stimulation. The P2 amplitude was larger with natural stimulation. The SD showed significantly larger amplitude for P2 and N1-P2 with /pa/. The GS showed shorter latency for P2, N2, and N1-P2 and larger amplitude for N2 with female speakers. The GL showed shorter latency for N2 and N1-P2 and larger amplitude for N2 with female listeners. **Conclusions**: Although several variables showed significance for N2, P2, and N1-P2, P1 and N1 did not show any significance for any variables. N2 and P2 of CAEP seemed affected by endogenous factors.

J Audiol Otol 2022;26(2):68-75

Keywords: Auditory evoked response; Natural speech response; Synthetic speech response.

Introduction

Auditory evoked potential (AEP) is an electrophysiological response stimulated by sounds. Cortical auditory evoked potential (CAEP), which belongs to the late response of AEP, is comprised of P1, N1, P2, N2, P300, and mismatch negativity (MMN) [1]. The P1 in the CAEP waveform reflects the activity in the secondary auditory cortex and Heschl's gyrus and has been utilized as an indicator of the maturity of the central auditory pathway in many studies. The N1, the first negative voltage component of CAEP, receives a contribution from the primary auditory cortex resulting in left and right hemispheres including intra- and inter-hemispheric activities. It is also known to be the most reliable component and influenced by changes of stimuli sensitively. The P2 is a complex response from the cerebral cortex and reflects the effectiveness of training in stimuli discrimination. The N2 is also a complex response from the superior auditory pathway including the thalamus and is affected by cognitive levels such as attention. In addition, the N1-P2 complex response can be used in the measurement of hearing sensitivity and neural processing of speech sound [2].

As speech sound stimulation is usually specified by the characteristics of consonants, the classification of the consonants is important for analyzing the response. Usually, consonants are classified as the articulation manner and place. The better accuracy was observed in perception with the plo-

sive and affricative consonants in Korean [3].

For the stimulation of CAEP, speech signals including natural and synthetic speeches were known to be effective as they were more related to cognitive processing [4]. It was noted that approximately 9 ms for P1 and N1 and 5 ms for P2 were faster with natural /a/ than synthetic /a/. And the natural speech sounds were more effective in recognition for both quiet or noisy environments than synthetically produced speech sounds [5] as they contained various real metrical characteristics such as intonation and stress which contributed to the sound accuracy for recognition [6,7]. However, when compared to click, tones and synthesized speech sounds, the natural speech sounds were thought to be less reliable for recording CAEP since they were highly complex time-varying signals. Also, synthetical speech sounds could be controlled and adjusted the acoustic features such as voice onset time (VOT) [8,9].

According to non-pathologic subject factors such as gender, CAEP waveforms showed the difference. One study reported a larger amplitude response in female than males for N1-P2 complex with tonal stimuli to both ears [10]. Also, the latency of P2 was significantly shorter when evoked with speech stimulation in female. But the amplitudes of P1, N1, and P2 were not differed by gender effect at the same study [11]. In another study, the latency and amplitude of N1 and P2 did not show any difference according to gender [12].

Several synthesizers have been introduced to facilitate the production of synthetic speech sounds including Microsoft's speech platform (https://www.microsoft.com), Klatt synthesizer (https://www.asel.udel.edu/speech/tutorials/synthesis/KlattSynth/index.htm), eSpeak (Hewlett-Packard, https://sourceforge.net/). Klatt synthesizer, developed in 1980, produced speech sound in a way of allocating particular formant frequency to phoneme but showed the poor quality of sounds [13]. Also, developed by Microsoft in 2011, Microsoft's speech platform was introduced as a standard speech synthesizer software. Based on Klatt synthesizer (1995), eSpeak was further developed following the Microsoft speech platform with an improvement of sound quality. It could also convert text-to-speech with Windows and Linux operating systems. Moreover, eSpeak has advanced accessibility providing language support systems over 43 languages depending on gender including the Korean language [14].

CAEP can be described as either endogenous or exogenous responses. P300 and MMN which are often called cortical P300 and MMN elicited by the oddball paradigm are categorized as endogenous among CAEPs. Generally, P1, N1, and P2 are reported as an exogenous response, characterized by external factors such as acoustic features of stimuli [15]. However, it is not clear whether N2 includes exogenous or endogenous factors [16]. CAEP waveforms showed the difference according to many non-pathologic subject factors. Although it has been rarely investigated to demonstrate gender effects, the amplitude of the N1-P2 complex was increased when tonal stimulation was presented to female listeners [10]. And the latency of P2 was decreased when speech stimulation was presented to female listeners [10]. The latencies and amplitudes of N1 and P2 latencies increased and amplitudes decreased as the age increased in adults [11,17].

As mentioned above, speech stimulation is quite effective in extracting the response of CAEP. When compared the difference between natural and synthesized vowels, the natural vowel sound showed a shorter latency. However, considerating the acoustic characteristics of Korean consonant and vowel(CV), CAEP studies were limited. Therefore, this study was pursued to understand CAEP waveforms with the Korean CV monosyllables and to utilize the findings clinically. Specifically, the CAEP characteristics were explored according to natural and synthetic speech stimulations, female and male speakers, and listeners using Korean CV composition. Two experiments were composed. The first experiment was to investigate and select the appropriate speech material which would elicit the most contrasting responses in perception among plosives from natural and synthetic speech sounds. The second experiment was to identify CAEP responses by asking the following four research questions: 1) Whether and how natural and synthetic speech sounds would affect the latency and amplitude of CAEP waveform; 2) Whether and how the selected monosyllables would change latency and amplitude of CAEP waveforms; 3) Whether and how latency and amplitude of CAEP waveforms would be affected by the gender of the speaker (GS); and 4) Whether and how the latency and amplitude of the CAEP waveform would be affected by the gender of the listener (GL).

Subjects and Methods

Participants

For comparison of natural and synthetic stimuli 21 young announcers (9 males, 12 females) participated. The mean age was 22 (standard deviation: ±1.7) years old. For CAEP measurement, 40 young adults (20 males, 20 females) participated. The mean age was 23.5 (standard deviation: ±2.04) years old. They all signed an agreement of research participation and were native Korean speakers with normal hearing thresholds better than 15 dB HL. The study was approved by the Institutional Review Board of Hallym University (HIRB-2019-071). All the participants provided written informed consent forms before initiation of the study.

Experimental material and equipment

The measurement was performed with GSI 61 (Grason-Stadler, Eden Prairie, MN, USA) and TDH-50P headphones in a soundproof room. All the participants showed A type of tympanogram with GSI 38 Autotymp (Grason-Stadler) and reported no history of otologic disease.

Concerning the stability of articulation manner, the plosive consonants /g/ and /b/ and aspirated sound of plosive consonants /k/ and /p/ were selected. When they were combined with the Korean vowel /a/, formulating four monosyllables, /ga/, /ba/, /ka/, and /pa/ were formulated. Then the monosyllables were recorded as speech sound stimulation. This natural speech sound stimulation was recorded from the Korean native professional male and female announcers using a microphone (RODE NR1A, Silverwater, NSW, Australia) in a soundproof room. For the synthetic speech sound stimulation, the speech sounds material for this study was recorded from eSpeak, which provided the Korean speech sound source at the homepage. Finally, natural and synthetic sound sourcies were prepared. With this Korean speech stimulation source, the six bisyllables were formulated in two presenting order (PO) tentatively, /ga/-/ka/, /ga/-/ba/, /ga/-/pa/, /ka/-/ba/, /ka/-/pa/, and /ba/-/pa/ as forwarding order, and /ka/-/ga/, /ba/-/ga/, /pa/-/ga/, /ba/-/ka/, /pa/-/ka/, and /pa/-/ba/ as backwarding order. These bisyllables were recorded in two stimulation modes (SM), natural and synthetic mode according to the gender of the speaker (GS). The stimuli were analyzed by Praat (versions of Microsoft Windows XP, Paul Boersma and David Weenink of the University of Amsterdam, USA) [18] and Computerized Speech Lab (CSL 4150B, KAYPENTAX Corp., Lincoln Park, NJ, USA). The audio files were edited to 48 kHz of sampling rate, mono channel, 16 bits, and 500 ms of maximum length. All stimuli were equally set at -20 dB RMS and presented at the most comfortable level (MCL) with the one-second interval.

Experimental procedure

All the bisyllables were presented randomly to the participants. After listening to the speech sound, they were asked to pick one what they heard and identify whether the sound was natural or synthetic. The trial practices were given to every participant 3−4 times to get used to the stimulation. Twelve bisyllables, six forward and six backwards, were presented according to four conditions by gender composition: male/male, male/female, female/male, and female/female, with two SM conditions, natural and synthetic modes. Totally, 96 stimulation numbers (12×4×2) were presented to each participant. Utilizing a laptop computer (NT930QAA-K38A, SAMSUNG, Seoul, Korea), the presentation of stimuli was given. The responses were obtained by clicking the buttons of the response pad (RB-740, SuperLab version 5.0.5; Cedrus, San Pedro, CA, USA).

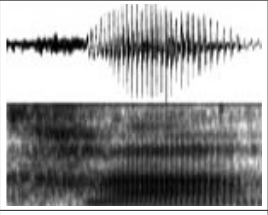
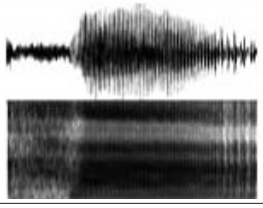
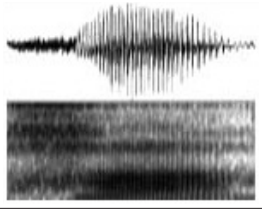
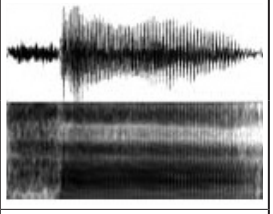
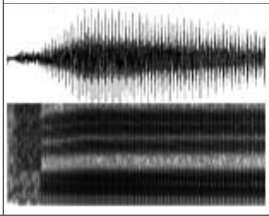
Waveform				
Frequency spectrum				
Natural speech	Male /ka/	Female /ka/	Male /pa/	Female /pa/
Pitch (Hz)	129	161	134	182
VOT (ms)	120	110	110	90
Formants (Hz)	F1=891, F2=1,435, F3=2,837	F1=1,016, F2=1,624, F3=2,667	F1=773, F2=1,493, F3=2,916	F1=1,040, F2=1,583, F3=2,617
Waveform				
Frequency spectrum				
Synthetic speech	Male /ka/	Female /ka/	Male /pa/	Female /pa/
Pitch (Hz)	89	177	89	170
VOT (ms)	80	80	60	50
Formants (Hz)	F1=754, F2=1,287, F3=2,715	F1=769, F2=1,339, F3=2,891	F1=727, F2=1,273, F3=2,663	F1=774, F2=1,354, F3=2,869

Fig. 1. Waveform, frequency spectrum, and other features of natural and synthetic sound /ka/ and /pa/. The top and bottom panels represent the natural and synthetic sounds respectively. VOT, voice onset time.

For the CAEP measurement, the monosyllables /ka/ and /pa/ were selected as the proper stimuli through the first experiment (Fig. 1). The stimuli were presented through insert earphone (ER3A, Etymotic Research, Elk Grove Village, IL, USA) to both ears simultaneously at 70 dB HL. Bio-logic (Natus Medical Incorporated, Mundelein, IL, USA) was utilized to obtain CAEP. Electrode impedance was maintained at less than 5 kΩ and inter-electrode impedance was maintained less than 1 kΩ. Rarefaction polarity stimuli were presented at a rate of 1.0 per second. Responses were filtered by a high-pass filter at 1 Hz and a low-pass filter at 30 Hz and averaged from 200 samples. Responses above 90 μV were rejected to minimize the artifact. After the subject was placed on the bed, stimuli were provided through insert earphone to both ears simultaneously. And the light was turned off to focus on the stimulus sound during the examination. The condition of the subject was continuously checked.

Statistical analysis

Statistical analysis was performed using SPSS statistical program (version 21, IBM Corp., Armonk, NY, USA). For experiment 1, four variables, SM, PO, GS, and GL were analyzed with ANOVA. The confidence interval of the result was set at $p<0.05$. Post hoc analysis was performed using a Bonferroni correction test. For experiment 2, the independent variables were SM and syllable difference (SD), between /ka/ and /pa/. The dependent variables were latencies and amplitudes of P1, N1, P2, N2, N1-P2 complex. Three-way mixed ANOVA was conducted for statistical analysis.

Results

Comparison of natural and synthetic stimuli

Only one main effect, SM, elicited the significant difference [F(1,20)=245.371, $p<0.01$] showing higher scores in the natural speech sound mode. The correction rates of the natural and synthetic speech sounds were 93% and 41%, respectively. Regarding the interaction effects, the better correction rates were observed with the condition of SM×PO in the natural speech sound for the forwarding order (93%), SM×GS in the natural speech sound produced by the male speaker (92%), PO×GS with the backwarding order of the female speaker (70%), and PO×GL with the forwarding order of the female listener (69%); SM×PO×GS in the natural speech sound with the backwarding order of the male speaker (94%), SM×GL×GS in the natural speech sound of the male listener and the female speaker (71%); and SM×PO×GS×GL in the natural speech sound with the backwarding order of the female listener and the male speaker (99%) (Table 1). Among all the stimuli, the correction rate difference was biggest (74%) at /ka/-/pa/ and /pa/-/ka/indicating that /ka/ and /pa/ evoked the most perceptual difference in terms of SM (Fig. 2). As a result, /ka/ and /pa/ were selected for the stimulus of the second experiment.

Table 1. Main and interaction effects of four variables, stimulation mode, presenting order, gender of listener, and gender of speaker

Condition	F	*p*
Main effect		
SM	245.371	<0.001
PO	0.006	0.938
GL	0.286	0.607
GS	0.396	0.532
Interaction effect		
SM×PO	16.155	<0.001
SM×GS	12.793	<0.001
SM×GL	0.196	0.997
PO×GS	6.005	<0.001
PO×GL	4.171	<0.001
GS×GL	1.304	0.236
SM×PO×GS	17.015	<0.001
SM×GL×GS	8.672	<0.001
PO×GS×GL	1.453	0.960
SM×PO×GS×GL	8.747	<0.001

SM, stimulation mode (natural versus synthetic speech sounds); PO, presenting order (forwarding versus backwarding orders); GL, gender of listener (male versus female listeners); GS, gender of speaker (male versus female speakers)

CAEP measurement

Depending on SM (natural and synthetic speech mode stimuli), SD (stimulus difference between /ka/ and /pa/), and GS, the average CAEP waveforms were depicted in Fig. 3. The latencies of P2 and N1-P2 complex were significantly shorter with the natural speech sound [F(1,78)= 7.723, $p<0.05$; F(1,78)=44.08, $p<0.05$, respectively] and the latency of N2 was significantly shorter with the synthetic speech sound [F(1,78)=7.723, $p<0.05$]. The amplitude of P2 was significantly larger with the natural speech sound [F(1,78)=24.95, $p<0.05$]. The amplitudes of P2 and N1-P2 for the /pa/ were significantly larger than the /ka/ [F(1,78)=5.679, $p<0.05$; F(1,78)=56.35, $p<0.05$; respectively], but the latency was not significantly different. The latencies of P2, N2, and N1-P2 complex of the female speaker were significantly shorter than those of the male speaker [F(1,78)=34.87, $p<0.05$; F(1,78)=17.88, $p<0.05$; F(1,78)= 8.491, $p<0.05$; respectively]. The amplitude of N2 was significantly increased with the female speaker [F(1,78)= 6.047, $p<0.05$]. The latencies of N2 and N1-P2 complex were significantly shorter with the female listener [F(1,36)=11.68, $p<0.05$; F(1,38)=4.970, $p<0.05$; respectively] and amplitude of

N2 was significantly larger with the female listener [$F(1,78)=6.047$, $p<0.05$] (Table 2, Fig. 4). There were significant differences in all CAEPs for interaction effects except N1. These significant differences were mostly identified from the interactions with two variables SM and GS. It was also notable that the interaction effects were consistent with the main effects except the effects of SD. To summarize, SM, GS, and GL were dominant factors which showed shorter latency and larger amplitude with the natural speech sound, the female speaker, and the female listener.

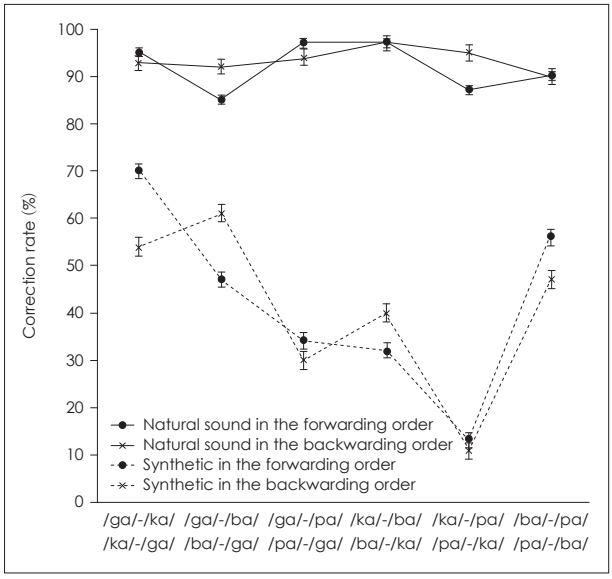


Fig. 2. Comparison of correction rates for the natural and synthetic speech sounds in the forwarding and the backwarding orders.

Discussion

The natural SM showed better perception ability than synthetic SM. The accuracy of the naturally spoken stimuli was also reported in the previous study with the intelligibility evaluation study of CV stimuli [7]. Lack of prosodic features and longer recognition time for synthetic speech sound were thought to be the reason [6]. Except for SM, the changes due to the condition of PO, GS, and GL were not significantly different. The forwarding and backwarding orders showed correction rates of 69% and 67% with no difference. However, the PO was reported to influence the results for both numeral and word stimuli in the literature [19]. When the characteristics of presentation orders of digit span were analyzed for an adult group whose auditory capability remained the same in the previous study. As a result, the forwarding digit span showed a 50% correction rate while the backwarding digit span showed only a 6% correction rate [19]. However, conventional findings of the order of presentation could not be applied to this study because the order of the stimuli of this study was assigned tentatively. The GS showed no effect on cognitive processing with many languages in the previous study [20]. This agreed with the result of this study. Neither GS nor GL did not show any difference in speech recognition (66% in male, 62% in female) and also the GL (66% in male, 67% in female).

The decreased latency and increased amplitude of P1 and N1 of the natural speech sound stimulation were not found in this study. Although the change of P1 and N1 were reported to be affected by exogenous factors such as duration and type of

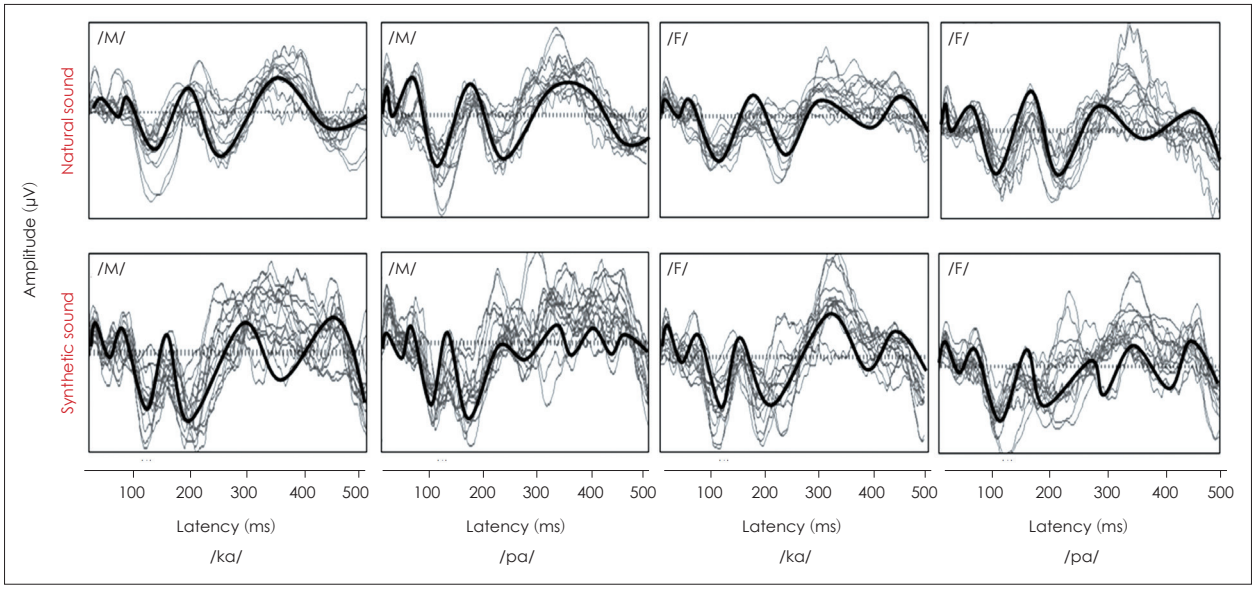


Fig. 3. Average waveforms obtained by the natural and synthetic sound /ka/ and /pa/ stimuli according to the gender of speakers were described in bold lines. The waveforms from the natural and synthetic sound stimuli were put at the top and bottom, respectively. M, male; F, female.

Table 2. Probability values of CAEP for main and interaction effects

CAEPs		Main effects				Interaction effects										
		SM	SD	GS	GL	SM× SD	SM× GS	SM× GL	SD× GS	SD× GL	GS× GL	SM× SD× GS	SM× SD× GL	SM× GS× GL	SD× GS× GL	SM× SD× GS×GL
P1	Latency	0.973	0.588	0.375	0.863	0.891	0.010*	0.104	0.012*	0.009*	0.676	0.658	0.814	0.070	0.017*	0.424
	Amplitude	0.293	0.801	0.850	0.714	0.459	0.913	0.042*	0.232	0.032*	0.005*	0.078	0.086	0.785	0.014*	0.286
N1	Latency	0.162	0.834	0.365	0.249	0.952	0.817	0.188	0.117	0.656	0.877	0.388	0.416	0.399	0.157	0.272
	Amplitude	0.323	0.499	0.163	0.678	0.288	0.567	0.750	0.532	0.731	0.885	0.279	0.537	0.691	0.922	0.068
P2	Latency	0.007*	0.153	<0.001*	0.112	0.042*	0.194	0.691	0.746	0.670	0.350	0.160	0.845	0.841	0.577	0.065
	Amplitude	<0.001*	0.020*	0.964	0.492	0.745	<0.001*	0.001*	0.078	0.013*	0.079	0.909	0.783	0.029*	0.344	0.873
N2	Latency	0.001*	0.065	0.002*	<0.001*	0.084	0.001*	0.106	0.001*	0.069	0.607	0.355	0.644	0.039*	0.030*	0.925
	Amplitude	0.337	0.109	0.016*	0.006*	0.165	0.017*	0.311	0.021*	0.016*	0.310	0.604	0.469	0.314	0.052	0.793
N1-P2	Latency	<0.001*	0.240	0.005*	0.032*	<0.001*	0.031*	0.188	0.009*	0.189	0.366	0.167	0.028*	0.524	0.467	0.060
	Amplitude	0.117	<0.001*	0.064	0.576	0.020*	0.010*	0.819	0.320	0.013*	0.045*	0.874	0.306	0.117	0.306	0.073

Greenhouse-Geisser value. *$p < 0.05$. CAEP, cortical anditory evoked potential; SM, stimulus mode; SD, syllable difference (/ka/ and /pa/); GS, gender of speaker; GL, gender of listener

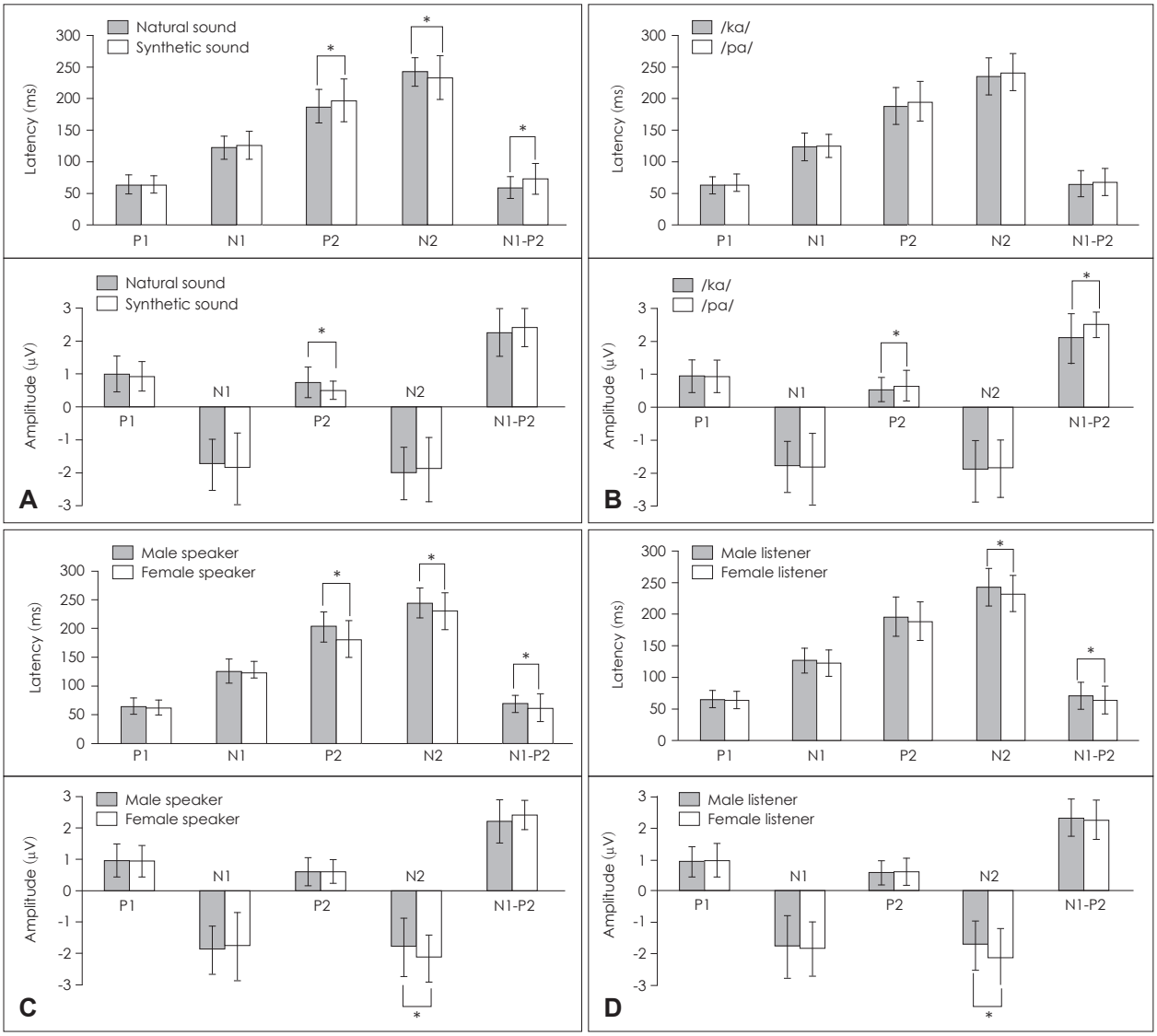


Fig. 4. Comparison of latencies (above) and amplitudes (below) according to stimulation mode between natural and synthetic sound (A), syllable difference between /ka/ and /pa/ (B), the gender of speaker (C), and the gender of listener (D). *$p < 0.05$.

stimulus and maturation of auditory system by the previous study [21], that phenomenon was not shown in our study.

The latency of P2, N2, and N1-P2 complex and amplitude of P2 showed significant differences between the natural and synthetic speech sound stimuli. The P2 and N2 were known to be useful to identify the change of neural processing depending on the speech stimulation as they were thought to be compound responses of the auditory cortex [22]. The first, second, and third formants were 915 Hz, 1,534 Hz, and 2,759 Hz for the natural speech stimulation, and 763 Hz, 1,313 Hz, and 2,784 Hz for the synthetic speech stimulation showing the notable differences in the stimulation materials of this study. It was thought that the acoustical characteristics including formant frequencies changed P2 and N2 responses as the previous study reported [23].

The statistical difference was elicited in the amplitudes of P2 and N1-P2 complex due to the SD. In the previous study, P2 which was involved in cognitive process and originate from the various area of the brain including auditory cortex was revealed to be advantageous in identifying changes of speech discriminating neural processing [22,24]. The statistical difference shown in this study might in the line with effect of cognitive processing of P2 shown in the previous studies. Also, N1 was affected by VOT and differences in duration [25]. However, the VOT duration of this study did not show a significant differences in /ka/ (98 ms) and /pa/ (78 ms) (Fig. 1).

The GS showed statistical significance in the latency of P2, N2, and N1-P2 complex. And the amplitude showed a statistical significance with N2. The previous study of gender difference in long-term average speech spectrum showed different spectrums according to the sex of speakers [26]. Every human has their own voice and different pitches within their vocal range [27]. In this study, the average speaker's pitch of the speech of males and females were 110 Hz and 172 Hz (Fig. 1). Also, the speech frequency according to the GS was an important factor influencing speech recognition [28]. When the voices of males and females were presented to the listeners who speak English as their own language, the average sentence recognition was 89.5% with female speakers, and 86.2% with the male speaker, showing a significant difference [29]. Conclusively, the significant difference in the latency of the P2 and N1-P2 complexes was due to the acoustic differences in the anatomical characteristics of the vocal organs between females and males.

The GL showed statistical significance in the latencies of the N2 and N1-P2 complex. And the amplitude showed statistical significance in N2. CAEP difference according to the GL was reported a different speech processing strategies for males and females depending to brain lateralization [30]. In general, women have better verbal strategies than men, and women's temporal lobes showed more balanced activity than men in passive listening situations where the left and right hemispheres were used equally [31]. In addition, it was reported that women responded more sensitively to auditory cognitive abilities than men [32]. The gender difference of the listener in the N2 latency and amplitude due to the asymmetry of the language-related hemispheres was found in this study.

Overall, the latency and amplitude of P1 and N1 were not affected by any of the variables. The latencies of P2 and N1-P2 complex were significantly different according to the SM and SD. The N2 latency and amplitude seemed to be affected by differences in perception, including attention and language perception in the cerebrum. Conclusively, when analyzing the results in the view of endogenous and exogenous features, it was found that N2 and P2 were mainly affected by the endogenous factor.

The study result showed the possibility of utilizing synthetic speech sound clinically, as the synthetic speech sound can be acquired by easy access to eSpeak and can be controlled more freely. However, only plosive /ga/, /ba/, /pa/, and /ka/ were used in the present study. In the future study, various syllables should be included to reveal a more general effect between the natural and synthetic speech sound signals. Therefore differences in CAEP waveforms should be investigated with not only the selected monosyllables, /ka/ and /pa/, but also with various meaningful monosyllables used in real life, as we use a great number of different syllables in various situations.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government Ministry of Science and ICT (No. 2020R1F1A1069236).

Conflicts of interest

The authors have no financial conflicts of interest.

Author Contributions

Conceptualization: Jinsook Kim. Data curation: Eunsung Lee, Saea Kim, Chanbeom Kwak. Formal analysis: Eunsung Lee, Saea Kim, Chanbeom Kwak. Investigation: Eunsung Lee, Saea Kim, Chanbeom Kwak. Methodology: Jinsook Kim, Woojae Han. Project administration: Jinsook Kim. Supervision: Jinsook Kim, Woojae Han. Validation: Woojae Han. Visualization: Jinsook Kim, Hyunwook Song. Writing—original draft: Hyunwook Song, Yerim Shin, Seungik Jeon, Eunsung Lee. Writing—review & editing: Hyunwook Song, Yerim Shin, Seungik Jeon, Eunsung Lee. Approval of final manuscript: all authors.

ORCID iDs

Hyunwook Song	https://orcid.org/0000-0003-4932-7330
Seungik Jeon	https://orcid.org/0000-0003-2288-0750
Yerim Shin	https://orcid.org/0000-0002-2914-6711
Woojae Han	https://orcid.org/0000-0003-1623-9676
Saea Kim	https://orcid.org/0000-0003-4233-8213

Chanbeom Kwak https://orcid.org/0000-0001-5657-7536
Eunsung Lee https://orcid.org/0000-0001-6525-3533
Jinsook Kim https://orcid.org/0000-0003-3440-2393

REFERENCES

1) Hall III JW. eHandbook of auditory evoked responses: principles, procedures & protocols. Pretoria: Pearson;2015. p.1-1749.
2) Ponton CW, Eggermont JJ, Kwong B, Don M. Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. Clin Neurophysiol 2000;111:220-36.
3) Choi SW. Perception and production of Korean lax, tense, and aspirated consonants: focused on Chinese learners. Studies in Linguistics 2018;47:21-36.
4) 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.
5) Reynolds M, Jefferson L. Natural and synthetic speech comprehension: comparison of children from two age groups. Augment Altern Comm 1999;15:174-82.
6) Cutler A, Dahan D, van Donselaar W. Prosody in the comprehension of spoken language: a literature review. Lang Speech 1997;40:141-201.
7) Raphael LJ, Borden GJ, Harris KS. Speech science primer: physiology, acoustics, and perception of speech. Philadelphia: Lippincott Williams & Wilkins;2007. p.1-346.
8) Tremblay KL, Friesen L, Martin BA, Wright R. Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. Ear Hear 2003;24:225-32.
9) Shaywitz SE. Overcoming dyslexia: a new and complete science-based program for reading problems at any level. New York: Vintage Books;2003. p.1-432.
10) Onishi S, Davis H. Effects of duration and rise time of tone bursts on evoked V potentials. J Acoust Soc Am 1968;44:582-91.
11) Swink S, Stuart A. Auditory long latency responses to tonal and speech stimuli. J Speech Lang Hear Res 2012;55:447-59.
12) Gölgeli A, Süer C, Ozesmi C, Dolu N, Aşcioğlu M, Sahin O. The effect of sex differences on event-related potentials in young adults. Int J Neurosci 1999;99:69-77.
13) Klatt DH. Software for a cascade/parallel formant synthesizer. J Acoust Soc Am 1980;67:971-95.
14) Duddington J. eSpeak. Sourceforge [Internet]. Sourceforge; 2012 [cited 2016 Nov 1]. Available from: URL: http://espeak.sourceforge.net/.
15) Wunderlich JL, Cone-Wesson BK, Shepherd R. Maturation of the cortical auditory evoked potential in infants and young children. Hear Res 2006;212:185-202.
16) Ritter W, Simson R, Vaughan HG Jr. Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. Psychophysiology 1983;20:168-79.
17) Rufener KS, Liem F, Meyer M. Age-related differences in auditory evoked potentials as a function of task modulation during speech-nonspeech processing. Brain Behav 2014;4:21-8.
18) Boersma P, van Heuven V. Speak and unspeak with PRAAT. Glot International 2001;5:341-5.
19) Swanson HL. Individual differences in working memory: a model testing and subgroup analysis of learning-disabled and skilled readers. Intelligence 1993;17:285-332.
20) Cambron NK, Wilson RH, Shanks JE. Spondaic word detection and recognition functions for female and male speakers. Ear Hear 1991; 12:64-70.
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) Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J Neurosci 2003;23:5545-52.
23) Näätänen R, Simpson M, Loveless NE. Stimulus deviance and evoked potentials. Biol Psychol 1982;14:53-98.
24) Tremblay KL, Shahin AJ, Picton T, Ross B. Auditory training alters the physiological detection of stimulus-specific cues in humans. Clin Neurophysiol 2009;120:128-35.
25) Burger M, Hoppe U, Lohscheller J, Eysholdt U, Döllinger M. The influence of temporal stimulus changes on speech-evoked potentials revealed by approximations of tone-evoked waveforms. Ear Hear 2009; 30:16-22.
26) Noh H, Lee DH. Cross-language identification of long-term average speech spectra in Korean and English: toward a better understanding of the quantitative difference between two languages. Ear Hear 2012; 33:441-3.
27) Chen SH. The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects. J Acoust Soc Am 2005;117: 3225-30.
28) Kim EO, Lim D. Effects of word difficulty and talkers on monosyllabic word recognition tests. Audiol 2006;2:102-6.
29) Bradlow AR, Torretta GM, Pisoni DB. Intelligibility of normal speech I: global and fine-grained acoustic-phonetic talker characteristics. Speech Commun 1996;20:255-72.
30) Cowell PE, Turetsky BI, Gur RC, Grossman RI, Shtasel DL, Gur RE. Sex differences in aging of the human frontal and temporal lobes. J Neurosci 1994;14:4748-55.
31) Phillips MD, Lowe MJ, Lurito JT, Dzemidzic M, Mathews VP. Temporal lobe activation demonstrates sex-based differences during passive listening. Radiology 2001;220:202-7.
32) Jang HJ, Kyung BP, Lee DL, Lee WB, Ryu SH. Considering elements of game design based on leaner's gender. KOCON 2011;11:128-36.