2012年 10月 13日，日本音響学会聴覚研究会資料，Vol. 42, No.7, H-2012-107
2012年10月13日，日本音響学会音響教育研究会資料，Vol. 7, No. 1, EDU-2012-11

Effects of Surrounding Contexts on English /ra/-/la/ Perception
－ For Educational Purposes －

Kanako TOMARU† and Takayuki ARAI‡

†‡Faculty of Science and Technology, Sophia University 7-1 Kioi-cho,
Chiyoda-ku, Tokyo, 102-8554 Japan
E-mail: †himawari.kanako@gmail.com, ‡arai@sophia.ac.jp

Abstract The present study reports perception of /ra/-/la/ continuum by native speakers of English and native speakers of Japanese in different contexts, i.e. speech and non-speech. In the speech context, a target syllable was presented as part of a sentence. In the non-speech context, a target was presented in-between pure tones. For native speakers of English, discrimination peak was observed in the non-speech context, but not in the speech context. That is, even native speakers of English had difficulties discriminating /ra/-/la/ continuum in the speech context. Results of native speakers of Japanese did not show any peaks in both context; however, experienced listeners had a kind of discrimination peak in the non-speech context. Through the experiment, we conclude that the type of context has effects on English listeners' discrimination of /r/-/l/ contrast. In addition, the study suggests that non-speech context may facilitate learners' discrimination of English /ra/-/la/ contrast.

Keywords speech perception, education, synthesized speech, /ra/-/la/ continuum, speech/non-speech contexts

1. Introduction

Conversational natural speech mostly comes in with sequence of speech segments to form phrases and sentences. Thus, putting grammatical and lexical problems aside, it is important not only to acquire phonemic contrast, but also to be able to recognize the contrast in a string of speech in order to understand spoken language. In the first language (L1) acquisition, children easily accomplish this goal by just being in their language environment. In the second language (L2) learning, on the other hand, perception of phonemes in ongoing speech often becomes a huge barrier against learners. Therefore, one of the final goals of English education is to improve learners' ability to recognize English phonemic contrast in any contexts, i.e. words, phrases or sentences.

Perception of English /r/-/l/ contrast by Japanese-speaking listeners is one of the strongest and the most famous pieces of evidence that L2 phoneme leaning is considerably difficult (for example [1-4, 6]). Japanese listeners perceive English /r/-/l/ contrast non-categorically whereas native speakers of English perceive it categorically (see [5] for categorical perception). Because Japanese does not have /r/ and /l/ as phonemes in its phonological inventory, native speakers of Japanese have considerable difficulty in perceiving English /r/-/l/ contrast [4].

A number of researchers have been seeking a way to train Japanese learners of English to learn /r/-/l/ contrast effectively 1) by using various talkers' utterance for practice [2, 3, 6], and 2) by emphasizing F3 difference in synthesized speech [7], for example. Our direction is also toward finding effective training on perception of /r/-/l/ contrast for native Japanese speakers. Especially, our main focus is on how to improve learners' ability to recognize the contrast in variety of contexts, i.e. words, phrases, and sentences. The present research is intended to investigate how surrounding sounds would influence phoneme perception. Surrounding sounds here are considered as context that phonemes are put in. In the present research, we had two types of sounds as surrounding context, i.e. speech and non-speech. We attempted to reveal if such types of surrounding context affect perception of English /r/-/l/ contrast. We employed relatively long speech/non-speech context in order to reveal 1) whether context type affects perception of native speakers of English, and 2) whether either context would facilitate Japanese listeners' perception of /ra/-/la/ contrast. Through the experiment, we revealed that the type of context had effects on English listeners' discrimination of /r/-/l/ contrast. In addition, the results suggest that L2 phonemic contrast may be effectively acquired within non-speech context.

2. Materials

2.1. Synthesis of /ra/-/la/ continuum

English /ra/-/la/ continuum was synthesized by using

Proc. Auditory Res. Meeting, The Acoustical Society of Japan, Vol. 42, No.7, H-2012-107, October 13, 2012
Proc. Education in Acoustics, The Acoustical Society of Japan, Vol. 7, No. 1, EDU-2012-11, October 13, 2012

Klatt cascade formant synthesizer [8]. The continuum consisted of ten stimuli (Step 1 to Step 10) following MacKain *et al.* [1]. Figure 1 provides schematic representation of trajectories of the first five formant frequencies. All synthesized syllables were 350 ms-long with 100-ms rising and falling periods before and after /ra/-/la/ part. So, Figure 1 shows trajectories without the rising or the falling period. The rising and the falling periods are also shown as gaps in Figure 2, which illustrates stimuli with surrounding contexts. See section 2.2 for detailed explanation of Figure 2.

Steady-state values of the first three formants, i.e., F1, F2 and F3, after transition were obtained from the naturally spoken /ʌ/ in "pronunciation" of a sentence "Clear pronunciation is appreciated" recorded by a male native speaker of English for TIMIT corpus [9]. The same utterance of the same speaker was employed for speech context (see Section 2.2.2). Table 1 shows the speaker's first three formant frequencies. Onset frequencies of F3 transition varied from 1609 Hz to 2827 Hz in nearly equal ten steps (Figure 1). Onset frequency of F1 and F2 were 376 Hz and 1299 Hz respectively. In addition to F2 transition, F1 steady-state and transitional duration were fixed throughout the continuum. F1 steady-state and transitional duration were 35 ms and 30 ms, respectively. F4 and F5 were 3250 Hz and 3700 Hz, respectively. The values were default of the synthesizer, and were fixed throughout the continuum. See Figure 1 for details.

2.2. Stimuli

2.2.1. Non-speech context

Surrounding contexts we had were speech and non-speech contexts. In the non-speech context, synthesized /ra/-/la/ syllables were placed in-between 1-kHz pure tones (Figure 2-a). The pure tone took 50 ms to fall before the onset of a syllable, and it took 50 ms to rise after the offset of a syllable. The synthesized syllables used for the non-speech context were identical to the ones used in the speech context. All synthesized syllables were 350 ms-long, and were preceded by 470 ms-long pure tone, and followed by 790 ms-long pure tone. Therefore, a whole stimulus with non-speech context lasted for 1610 ms (Figure 2). Ten stimuli were newly created for non-speech context because we had ten synthesized syllables.

2.2.2. Speech context

In the speech context, synthesized English /ra/-/la/ syllables appeared as part of a naturally spoken sentence. The sentence employed in the current experiment was:

Table 1. Values of the first three format frequencies of /ʌ/ uttered by a male native speaker of English from the TIMIT corpus (Hz).

F1	670
F2	1357
F3	2788

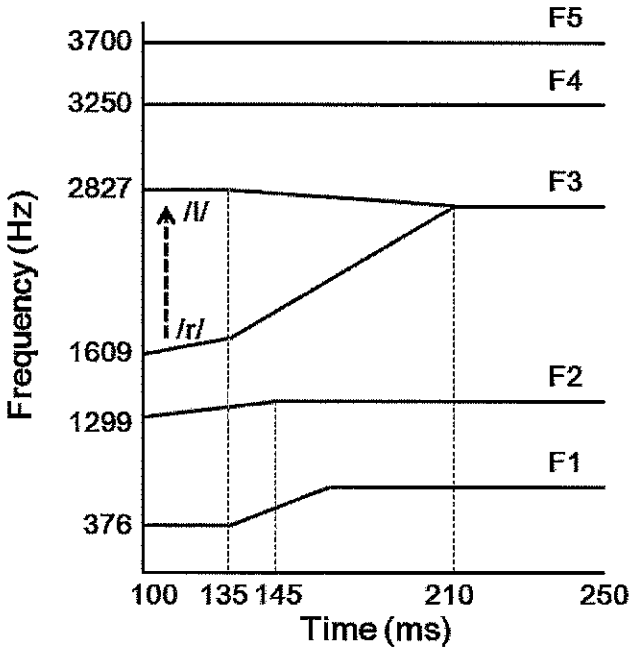

Figure 1. Schematic representation of trajectories of format frequencies from F1 to F5. The figure shows trajectories after 100-ms rising period.

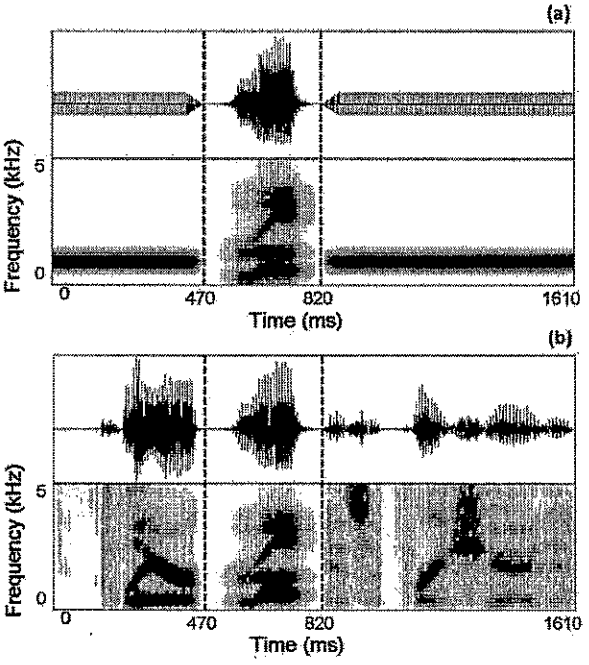

Figure 2. Examples of the spectrograms of stimuli for non-speech context (a), and speech context (b).

"Clear pronunciation is appreciated." The sentence was uttered by a male native speaker of English for TIMIT corpus [9], who was the same speaker we employed for /ra/-/la/ continuum synthesis (see Section 2.1). Synthesized syllables were replaced with the word "pronunciation" in the sentence. The syllables employed for the speech context were identical to those employed for the non-speech context. The newly created sentences looked like the one shown in Figure 2-b. All synthesized syllables were 350 ms-long, and were preceded by 470 ms-long speech sound, and followed by 790 ms-long speech sound. Therefore, a whole stimulus lasted for 1610 ms, which exactly matched to the length of stimuli in non-speech context (Figure 2). We, again, had ten newly created stimuli for speech context.

3. Experiment

3.1. Listeners

3.1.1. Native speakers of English

Nine native speakers of English (5 males, 4 females) participated in the experiment. Eight of them were from the United States, and one of them was from United Kingdom. Ages ranged from 20 to 29 years (mean 21.9 years). None of them reported any hearing problems at the time of the experiment. All listeners were recruited at Sophia University. They have resided in Japan for 3 months to 11 months at the time of the experiment (mean 5 months).

3.1.2. Native speakers of Japanese

Fifteen native speakers of Japanese (5 males, 10 females) participated in the experiment. Ages ranged from 18 to 44 years (mean 23.3 years). All listeners were either undergraduate or graduate students of Sophia University at the time of the experiment. All listeners had English classes in middle-high, high and undergraduate schools in Japan. Two listeners also had English instructions in elementary school in Japan. Listeners had no experience of residing in English-speaking countries except for five listeners: Listener 1, Listener 10, and Listener 14 have resided in the United States for four months, one year, and 2 weeks, respectively; Listener 12 has resided in Canada for 1 month; Listener 15 has resided in Australia for 2 months. Listener 14 also had experience residing in the United Kingdom for 3 weeks. Fourteen listeners had scores or grades on at least one of the following tests on English competence: Test of English for International Communication (TOEIC®), Test of English a Foreign Language (TOEFL®), and Test in Practical English Proficiency (EIKEN). None of them reported any hearing problems at the time of the experiment.

3.2. Procedures

3.2.1. Instructions

We employed AXB discrimination paradigm. Both English and Japanese listeners discriminated synthesized syllables in speech and non-speech contexts. The syllables were paired such that each pair (AB) differed by two steps in the continuum, i.e. Step 1－3, Step 2－4, Step 3－5, Step 4－6, Step 5－7, Step 6－8, Step 7－9, and Step 8－10.

All listeners had experiment on the non-speech context first. For syllables in the non-speech context, the following instructions were given to each group of listeners. English instruction was given to English listeners, and Japanese instruction was given to Japanese listeners: 1) You will hear a long sound with a syllable in the middle three times. Concentrate on the syllables in the middle, and say if the second syllable is more similar to the first, or to the third, and 2) このセッションでは、ある音節がピーという音に挟まれて3つずつ流れます。その音節を注意深く聞いて、2つ目に聞いた音節が、1つ目あるいは3つ目の、どちらと同じだったか答えてください。

For syllables in the speech context, the following instructions were given to each group of listeners. English instruction was given to English listeners, and Japanese instruction was given to Japanese listeners: 1) You will hear a man read a sentence, three times. Sentences you will hear will be like this: *Clear xx is appreciated*. Please concentrate on the *xx* part of the sentence, and say whether the second sentence sounds more similar to the first, or to the third, 2) 男の人が、*Clear xx is appreciated* という文章を3回読み上げます。*xx* の部分には、ある音節が入ります。その音節に注目して、2回目に聞いた文章に出てきた音節と同じ音節が含まれているのは、1回目あるいは3回目のどちらなのか、クリックして答えてください。

3.2.2. Experimental settings

We gave written instructions and experimental sessions to listeners by using Praat software [10]. The listeners made responses by clicking buttons that appeared on the screen. Paired stimuli that were 2-steps apart were presented as AAB, ABB, BAA, and BBA. There were three repetitions for each presentation, so listeners made 12 judgments for each pair. Thus, this makes total of 96 judgments for one session, i.e. speech or non-speech (8 pairs × 4 presentations × 3 repetitions = 96 judgments). AXB presentations were made randomly

within each session. For purposes of familiarizing the listeners with the tasks, a short practice session was held prior to the main experimental sessions. Stimuli were presented diotically via Sennheiser HDA 200 headphones at participants' comfortable listening level.

4. Results

4.1. Native speakers of English

Figure 3 shows average correct rate for English listeners. As you notice easily, the shape of the discrimination function for each context differs. When syllables were perceived in non-speech context, discrimination peak was obtained. The discrimination peak is a piece of evidence that shows categorical perception of English listeners [5]. When the same syllables were perceived in speech context, however, the peak was not obtained. These results suggest that English listeners' discrimination performance varies depending on the type of surrounding context.

4.2. Native speakers of Japanese

The discrimination function for Japanese listeners, on the other hand, did not show any specific peaks for both contexts. Figure 4 shows average correct rate for Japanese listeners. Having no discrimination peaks means that categorical perception was not obtained for Japanese listeners in both contexts. Such results are expected since neither /ra/ nor /la/ forms phonemic category in Japanese. In addition, correct rates of discrimination looks about the same in both speech and non-speech contexts. The results of the experiment, thus, seem to tell us that neither type of context facilitates Japanese listeners' perception of /ra/-/la/ contrast.

However, when we analyzed the data according to the listeners' experience of residing in English-speaking countries, we obtained some sign of the contextual effects on Japanese listeners' /ra/-/la/ discrimination. In the current experiment, five listeners had experience residing in English-speaking countries. The five listeners will be called experienced listeners, and the remaining listeners will be called non-experienced listeners, hereafter. Figure 5 shows the average correct rate of discrimination in non-speech context for experienced and non-experienced listeners. The function for the experienced listeners has a kind of a discrimination peak around the stimulus pair 6 −8. For the non-experienced listeners, on the other hand, the function does not show any peaks. Figure 6 shows the average correct rate in speech context. In the speech context, shapes of the discrimination function are rather flat for both experienced and non-experienced listeners.

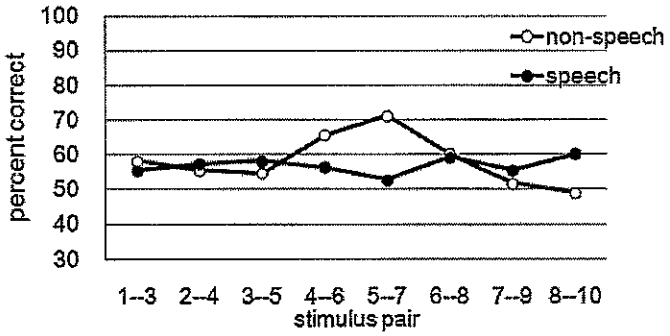

Figure 3. Average correct rate (%) of discrimination in non-speech context (open circle) and in speech context (closed circle) for native English speakers.

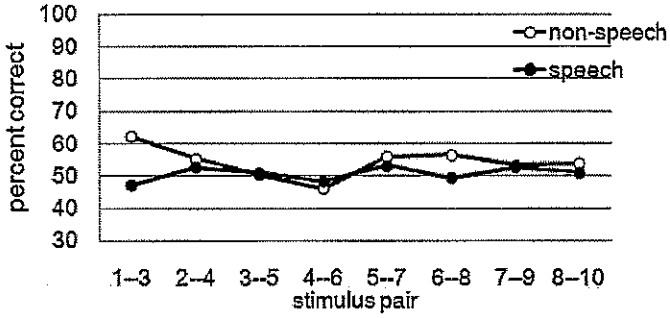

Figure 4. Average correct rate (%) of discrimination in non-speech context (open circle) and in speech context (closed circle) for native Japanese speakers.

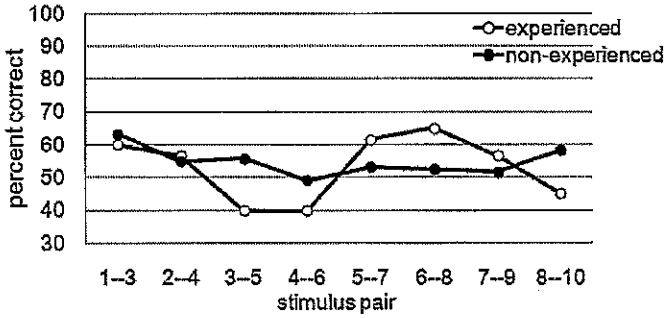

Figure 5. Average correct rate (%) of discrimination in non-speech context for experienced listeners (open circle) and non-experienced listeners (closed circle).

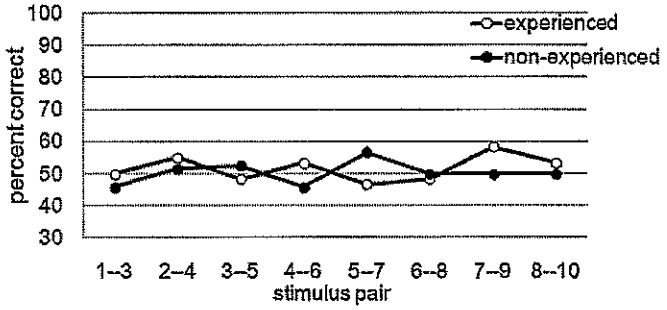

Figure 6. Average correct rate (%) of discrimination in speech context for experienced listeners (open circle) and non-experienced listeners (closed circle).

Especially, it is interesting that a peak of correct rate obtained in the non-speech context for the experienced listeners, was not observed in the speech context. The data may imply that surrounding context have some effects on experienced listeners rather than on non-experienced listeners.

5. Discussion

5.1. Native speakers of English

The purpose of the present research is to investigate the effects of surrounding context on speech perception. The current experiment revealed that the type of surrounding context had a great influence on perception of native speakers of English. That is, a discrimination peak was obtained for English listeners in non-speech context, but not in speech context. The questions here are: 1) would the type of surrounding context also affect the results of identification test? and 2) why did speech context influence English listeners' discrimination performance?

For the first question, it is unlikely that the type of surrounding context would affect identification test considering phonemic status of /ra/ and /la/ in English: native English speakers should show categorical perception in identification test with speech context. Therefore, the effects of surrounding context should be limited to discrimination task. In the current experiment, the length of non-speech context matched to that of speech context. Thus, the contextual effects observed here has nothing to do with the interval duration between syllables; rather, the type of context, i.e. speech or non-speech, had an influence on perception. Therefore, further experiments are needed to clarify what are the factors that make discrimination difficult in the speech context.

For the second question, we pick up following two possibilities among many: 1) naturalness of the sentence and 2) existence of complex acoustic information in speech. The sentence used for the current experiment, "Clear ra/la is appreciated", sounded rather unnatural in terms of its meaning. In daily conversations, listeners must rely not solely on acoustic information of the speech, but also on grammatical and lexical information to understand spoken sentences. However, the listeners may not have been able to use the strategies they use as usual in the current experimental settings because the stimulus sentence lacked naturalness. In addition, because speech contains a lot of information concurrently, it is possible that listeners' got confused about what is the specific cue they are to focus on in speech context. For non-speech pure tone, on the other hand, listeners do not need to have such concern because the pure tone consisted of only one band of frequency and did not have amplitude fluctuations. Thus, how naturalness and information complexity are involved in discrimination in speech context should be clarified in the further experiments.

5.2. Native speakers of Japanese

The effects of surrounding contexts on perception of /ra/-/la/ contrast were not obvious for native speakers of Japanese: no discrimination peak was obtained from average discrimination function. However, when analyzing the discrimination data in terms of experience of residing in English-speaking countries, discrimination function for experienced listeners changed its shape depending on the type of context: kind of a discrimination peak was observed in non-speech context for the experienced listeners. Four of the five experienced listeners, i.e. Listener 10, Listener 12, Listener 14, and Listener 15, reported that they used English for daily conversation at the time of living, and one remaining listener, i.e. Listener 1, had relatively high score and grade on TOELC® and EIKEN, i.e. 845 and Pre-1. Thus they may be accustomed to hear English sounds more than non-experienced listeners. Experience and high competence of English should be ones of the most effective factors that affect discrimination performance in the non-speech context. However, the number of experienced listeners for the current experiment was rather small, and their residing periods varied from two weeks to four months. In addition, the experienced listeners' levels of English competence also varied. Thus, in the further research, we will need to control the amount of experience of listeners in order to clarify if discrimination peak kind of rise of correct rate in non-speech context was due to facilitating effect of the context on experienced listeners.

5.3. Educational implications

As we mentioned in Introduction, English education should improve learners' ability to recognize English phonemic contrast in any speech contexts, i.e. words, phrases and sentences. Thus, learners of English have to be trained to perceive phonemic contrast within continuous flowing speech at some point. However, as revealed by the current experiment, even native speakers of English had difficulties discriminating /ra/-/la/ contrast in speech context. The results suggest that speech context affects discrimination performance of listeners even they are native speakers of English. Thus, phoneme

discrimination tasks in speech context may not be effective for training L2 learners to recognize phonemic contrast in continuous speech. However, there are some questions remained to be answered before jumping to conclusions. Two of the many are: is it really ineffective to train learners under speech context? and would it be effective to train Japanese learners of English under non-speech context, then? We are interested to clarify if there would be any contexts, either speech or non-speech, which would facilitate learners' discrimination ability of English phonemes. We will include investigation of effects of surrounding contexts on discrimination training of Japanese learners of English as a further issue.

6. Conclusion

The present research tried to reveal 1) whether context type, i.e. speech or non-speech, affects perception of native speakers of English, and 2) whether either context would facilitate Japanese listeners' perception of /ra/-/la/ contrast, through a perceptual experiment. The results of the experiment showed that English listeners perceived stimuli categorically in non-speech context, but not in speech context. Therefore, the experiment revealed that surrounding context influenced perception of native speakers of English. In addition, we showed that non-speech context may have encouraged experienced Japanese learners' categorical perception of English /ra/-/la/ contrast. We will conduct further experiments to clarify the effects of surrounding context on L2 perception.

References

[1] K. S. MacKain, C. T. Best, and W. Strange, "Categorical perception of English /r/ and /l/ by Japanese bilinguals," Applied Psycholinguistics, vol.2, no.4, pp. 369-390, 1981.

[2] W. Strange, and S. Dittman, "Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English," Perception & Psychophysics, vol.36, no.2, pp. 131-145, 1984.

[3] S. E. Lively, D. B. Pisoni, R. A. Yamada, Y. Tohkura, and T. Yamada, "Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories," Journal of the Acoustical Society of America, vol.96, no.4, pp. 2076-2087, 1994.

[4] R. A. Yamada, and Y. Tohkura, Perception of American English /r/ and /l/ by native speakers of Japanese, in Speech Perception, Production and Linguistic Structure, eds. Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka, pp. 155-174, IOS Press, Washington, 1992.

[5] A. M. Liberman, K. S. Harris, and H. S. Hoffman, "The discrimination of speech sounds whining and across phonetic boundaries," Journal of Experimental Psychology, vol.54, pp. 358-368, 1957.

[6] J. S. Logan, S. E. Lively, and D. B. Pisoni, "Training Japanese listeners to identify English /r/ and /l/: a first report," Journal of the Acoustical Society of America, vol.89, no.2, pp. 874-886, 1991.

[7] R. Danou, S. Iwamiya, and T. Furuhashi, "A study on training method to distinguish between L and R by enhancing frequency features of L/R," Proc. Spring Meeting, ASJ, pp. 523-526, 2012.

[8] D. H. Klatt, and L. C. Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," Journal of the Acoustical Society of America, vol.87, no.2, pp. 820-857, 1990.

[9] V. Zue, S. Seneff, and J. Glass, "Speech database development at MIT: TIMIT and beyond," Speech Communication, vol.9, no.4, pp. 351-356, 1990.

[10] B. Paul, and W. David, "Praat: doing phonetics by computer [Computer program]", ver.5.3.23, retrieved 7 August 2012 from http://www.praat.org/