Compensatory Speech Response to Time-Scale Altered Auditory Feedback

Rintaro Ogane¹, Masaaki Honda²

¹Graduate School of Sport Sciences, Waseda University, Japan
²Faculty of Sport Sciences, Waseda University, Japan

r.ogane@ruri.waseda.jp, hon@waseda.jp

Abstract

In this experiment, we examined speech compensation to timescale altered auditory feedback during the transition of the semivowel for a target utterance of /ija/. We examined speech compensation and the after effect in terms of three acoustic features, the maximum velocities on the F1 and F2 trajectories (VF1 and VF2) and the F1-F2 onset time difference (TD) during the transition, as well as listening test on the feedback speech. As the result, speech compensation was observed in VF1, VF2 and TD. The magnitude of speech compensation in VF1 and TD monotonically increased as amount of the time-scale perturbation. Speech compensation for time-scale altered auditory feedback is carried out primarily by changing VF1 and secondarily by adjusting VF2 and TD. Also, it is activated mainly by detecting the speed change in altered feedback speech sound.

Index Terms: altered auditory feedback, time-scale modification, after effect

1. Introduction

Speech production and sensory perception have a linkage in ongoing speech motor control as well as speech acquisition. Speech motor control employs various sensory information, including that of auditory feedback, tactile feedback and somatosensory feedback to achieve own utterance target. In particular, auditory feedback is significantly linked to speech production as well as speech acquisition. In order to investigate the linkage between speech perception and production, altered auditory feedback where the fundamental frequency or the formant frequencies of the uttered speech altered, has been widely used. Pitch-shifted auditory feedback [1–3] provided evidence of rapid compensatory response in the laryngeal control for maintaining the target pitch. Formant-shifted auditory feedback [4–7] also provided evidence of adaptation of the articulation to cancel the formant shift with relatively slow adaptation speed.

Most previous studies focused on the compensatory response to the stationary auditory perturbation. On the other hand, there have been few studies on speech compensation when non-stationary auditory feedback was applied during the utterance (e.g. [8]). In speech motor control, spectral dynamics is a significant feature for characterizing the consonants as well as semivowels. Dynamical characteristics of speech acoustics are also important perceptual features in identifying these phonemes.

In this study, we focus on speech compensation in response to non-stationary altered auditory feedback which is implemented by the local time-scale modification. We examine speech compensation and after effect on acoustic features in response to timescale altered auditory feedback. We also discuss how the speech compensation is activated by simply detecting the speed change in the feedback sound or by detecting the phonemic change caused by modification of the time-scale.

2. Methods

2.1. Preliminary perception test

A preliminary perception test was conducted to examine the perceptual phonemic change in the time-scale altered speech as the time-scale factor varies. Stimuli are the time-scale altered speech for the time-scale factors from 0.6 (shortened) to 2.5 (lengthened). The original speech was taken from an utterance of /ija/ spoken by a male subject. Ten listeners participated in this perception test, and they judged each stimulus as either /ija/ or /ia/. As the result, the subjects perceived stimulus as /ija/ from 0.6 to 1.2. On the other hand, the subjects perceived stimulus as /ia/ from 1.8 to 2.5. The interval between 1.2 and 1.8 is the boundary between /ija/ and /ia/. Therefore, we selected the time-scale factors from 1.0 to 1.8 for altered auditory feedback experiment.

2.2. Auditory feedback system

The altered auditory feedback system is illustrated in Fig. 1. Uttered speech was recorded using a microphone (ECM-672, SONY, Japan) and the speech signal was digitized at a sampling rate of 10 kHz after being low-pass filtered at 5 kHz (MS-521, NF Corporation, Japan). The signal was processed in a digital signal processor (TMS320C6713B, TEXAS INSTRUMENTS, USA) to alter the time-scale of the uttered speech using the TD-PSOLA method [9] in real time. The altered speech signal was then fed back into both ears of the subject through earphones (ER-4P, ETYMOTIC RESEARCH, USA) after low-pass filtered at 5 kHz. An EGG signal was simultaneously acquired for detecting the pitch marks on the speech waveform which was used in the TD-PSOLA processing. The pitch marks
were determined by detecting the instant when the time deriva-
tive EGG signal exceeded a chosen threshold. The onset time of the transition of semivowel, where altered auditory feedback was set at on, was automatically determined by detecting the instant when the squared sum of the delta LPC cepstrum co-
efficients obtained from the uttered speech signal exceeded a threshold. The perturbation interval was set at approximately 190 ms, which was adjusted to the utterance speed by the sub-
ject. The total time lag in the processing was 40 ms for an analy-
ysis window length of 10 ms, because the TD-PSOLA process-
ing needs the pitch mark in the next window block. The time lag is tolerated for the additional delayed auditory feedback ef-
fect [10].

The masking noise for masking the bone conduction speech was not mixed with the feedback speech sound, because the noise masks small changes in the altered feedback speech. The sound level at the earphones was in the range of 85 to 98 dB SPL. A guide tone signal, which was a series of four tones, was mixed with the feedback speech in order to synchronize the utterance timing with the guide tone. The first three pre-
tones were used for instructing the utterance start timing and the fourth guide tone was used for synchronizing the onset time of the semivowel utterance.

2.3. Subjects and procedure

Ten male subjects participated in the experiment. Subjects were native Japanese speakers and had no reported impairment of hearing or speech. They were instructed to speak the target di-
syllable /ija/ in synchronizing with guide tones mixed with the feedback speech. The experimental session was conducted for time-scale altered auditory feedback conditions with four time-
scale factors of 1.2, 1.4, 1.6, and 1.8. Time-scale modification was applied during the transition of the semivowel /j/ of the tar-
get disyllable /ija/ where the transition interval was lengthened according to the time-scale factor.

In each session, the subject repeated five utterance parts. Each utterance part was divided into three conditions, 10 control trials, 20 perturbed trials and 10 return trials. In perturbed trials, auditory feedback was changed in a step-wise manner. In the return trials, auditory feedback was changed in a step-wise manner to the normal auditory feedback condition. In each ut-
terance session, the subject uttered the target disyllable in a 2 sec length recording interval and had a 3 sec length pause be-
tween the successive utterances. Utterance data for a total of 800 trials was collected for each subject. Before the experi-
ment, the subjects trained to speak the utterance of the target disyllable synchronizing with the guide tones mixed with the feedback speech. In another training session, the subjects ut-
tered several times the target disyllable in both normal auditory feedback and altered auditory feedback with a time-scale factor of 1.6. Using these utterance data, we adjusted several control parameters in the experimental setup.

2.4. Data analysis

We examined three acoustic features which were obtained from the uttered speech. The formant frequencies were obtained for a 50-ms analysis window at every 1-ms frame period using Praat [11]. VF1 and VF2 is defined as a peak of ΔF(t) during the transition of the semivowel. TD is defined as the interval from the F2 onset time to the F1 onset time, F1 onset time minus F2 onset time, in the transitions of their trajectories. The F1 or F2 onset time was determined at the instant where each trajectory exceeds 20% of the extent of the formant change during the transition.

Each acoustic feature was normalized by subtracting the av-
average over ten control trials from the original value. Thus, each acoustic feature is relative value to the average over control tri-
als. Then, each acoustic feature was averaged over ten subjects and five repeated utterance trials to obtain the group data. The magnitude of speech compensation in each acoustic feature for every utterance trial was obtained by averaging the acoustic fea-
ture over the last five perturbed trials. Similarly, the magnitude of after effect was obtained by averaging the acoustic feature over the last five return trials. The difference of the magnitudes in each acoustic feature between the control condition and per-
turbed condition or return condition was examine by using t-
test.

3. Experimental results

3.1. Time course of speech compensation

Figure 2 shows a time course of speech compensation for three acoustic features. Every acoustic feature rapidly increases just after the beginning of perturbed trials and saturates in two to five trials. The variance each acoustic feature in the time course is small for VF1, TD and relatively large for VF2. As compared the time course of VF1 and VF2, the magnitude of speech compen-
sation in VF1 monotonically increases in order of the time scale factors from 1.2 to 1.8, but the difference of the magnitude in VF2 among the time scale factors is small. This indicates that speech compensation by the velocity change is more dominant in VF1 more than VF2. TD is also employed in speech compen-
sation. Here, the positive value of TD indicates that the onset timing of F2 during the transition of semivowel occurs earlier than that of F1. The time course in TD is more variant for the small time scale factor, in particular for 1.2. The magnitude of TD shows tendency to increase as the time scale factor changes from 1.2 to 1.8.

After the time-scale altered auditory feedback goes back to the normal condition, every acoustic feature shows a tendency to gradually decrease and approach to the value in the control condition. The time course of VF1 during return trials shows clear tendency to approach toward the value in the control con-
dition for every time scale factor. The slope of the time course in return trials is large as the time factor is large. At the final trial in the return condition, there remains difference of VF1 between the control condition and the return condition. The amount of after effect in VF1 is larger as time scale factor becomes large. Although the time course of VF2 during return trials shows ten-
dency similar to VF1, it is more variant than that of VF1 and the decreasing slope is gentle.

3.2. Magnitude and speed of speech compensation

The magnitude of VF1 in the perturbed condition monotononi-

cally increasing as the time-scale factor increases from 1.2 to 1.8 as shown in Fig. 3. This indicates that more effort is made for speech compensation by increasing the F1 velocity as amount of time scale perturbation increases. The difference of magnitudes between speech compensation and the control was statistically significant for every time scale factor of 1.2 to 1.8 (p<0.01). For the magnitude in the F2 velocity, the difference of magnitudes between speech compensation and the control was statistically significant for every time scale factor of 1.2 to 1.8 (p<0.01). However, as shown in Figs. 2 and 3, the magni-

tude in VF2 slightly increases as the time-scale factor changes from 1.2 to 1.4, but the magnitude saturates above 1.4. The
difference of the magnitude between every combination from four time scale factors was not statistically significant (p > 0.3). When the magnitudes of VF1 and VF2 in Fig. 3 are compared, the magnitude of VF1 is larger than that of VF2 for the time scale factors of 1.6 and 1.8, and the magnitude of VF2 is larger than that of VF1 for the time scale factor of 1.2 and 1.4. For the magnitude of the TD, there is an increasing tendency as the time-scale factor increases from 1.2 to 1.8, but there is not clear difference in TD among time scale factors of 1.4, 1.6 and 1.8. The difference in the magnitude of TD between the control and perturbed conditions is statistically significant for time scale factor 1.6 and 1.8 (p < 0.01). The magnitude in TD increases as the time-scale factor changes from 1.2 to 1.8. The difference of the magnitudes between the time scale factors of 1.2 and 1.8 was statistically significant (p < 0.01). The adaptation speed is characterized by the number of trials in the perturbed condition in which the time course saturates. As shown in Fig. 2, the time courses of VF1 and TD during the perturbed trials are less variant and it saturates in approximately three trials as every time-scale factor. On the other hand, the time course of VF2 during perturbed trials is more variant, it is difficult to explicitly determine the number of trials in which it saturates. Each time courses reaches in approximately five trials to the level of the magnitude averaged over the last five trials in perturbed trials.

3.3. Magnitude of after effects

The after effect caused by the time-scale altered feedback occurs in the first several trials in return trials and remains in VF1 and TD at the final 10th return trial for the large time scale factors (cf. Fig. 2). As shown in Fig. 4, the magnitude of after effect is statistically significant in VF1 for the time scale factor 1.8 (p < 0.01), in VF1 for 1.4 and 1.6 (p < 0.02), in VF2 for 1.4 (p < 0.01), in TD for 1.6 and 1.8 (p < 0.01).

3.4. Listening test

We performed a listening test to examine speech accuracy by the speech compensation. For this listening test, the 24 sound stimuli (5 control trials, 12 perturbed trials, 7 return trials) were selected from 40 trials in one utterance session for every time-scale factor.
scale factor and every subject. Ten listeners judged each stimulus as /ija/ or other. Figure 5 shows a result of listening test for feedback sound. The horizontal axis indicates the utterance trial number and the vertical axis indicates the identification score in percentage at which the stimulus is perceived as /ija/. Of note, the identification score drops at the first perturbed trial (11th) for every time-scale factor because the subject cannot predict the altered feedback in the first perturbed trial. For the perturbed conditions with the time-scale factors of 1.2 to 1.6, the score gradually increases during perturbed trials and approaches the score in the control trials in approximately five perturbed trials. However, for the perturbed condition with the time-scale factor of 1.8, the score increases in perturbed trials but it is still lower than that in the control trials at the final perturbed trial. In return trials from the 31st to 40th trials, the identification score shows small drop at the first return trial (31st) for every time-scale factor, and then rapidly improves and approaches to the score compatible to that in the control trials.

4. Discussions

The experimental results showed that the speech compensation in response to the time-scale altered auditory feedback occurs in VF1, VF2 and TD. Magnitudes of speech compensation in VF1 and TD are monotonically increasing as the time scale factor (amount of perturbation magnitude) increases, as while VF2 shows less relationship between amount of compensation and perturbation magnitudes. Speech compensation by increasing VF1 means that F1 changes with the speed proportional to the altered time scale in auditory feedback signal. On the other hand, speech compensation by increasing VF2 and TD means that F2 changes earlier to F1 with an increased speed in the transition of the semivowel in /ija/. This suggests that the speech compensation to time scale altered auditory feedback is mostly carried out by changing the speed of F1 and partially by adjusting the speed of F2 and the onset time of F2 in the transition.

The acoustical change of F1 and F2 in the utterance of the disyllable /ija/ is related to the jaw opening and tongue motion from the front to back positions during the speech articulation, respectively. Increasing VF1 and VF2 correspond to an increase in the motion speed of the jaw and the tongue. Also, an increase in TD indicates that the tongue motion occurs earlier than the jaw opening in compensatory response to the altered feedback. Thus, the changes in VF1, VF2 and TD suggest that the speech compensation is implemented not by increasing both jaw and tongue motion speeds simultaneously in proportion to amount of the perturbation magnitude but mostly by increasing the jaw opening speed and partially by moving the tongue with an increased constant speed and moving earlier the tongue in the transition.

Experimental results showed a significant change in the magnitude of speech compensation in VF1, VF2 and TD as compared to that in the control condition. Also, the magnitude in VF1 monotonically increased as the time-scale factor increased. There are two possible interpretations for speech compensation due to the time-scale altered auditory feedback. One is that speech compensation is activated by detecting the speed change. Another is that it is activated by detecting the phonemic change. The former interpretation predicts change in the magnitude of speech compensation when the time scale perturbation is applied. Also, if the motor control is adjusted according to amount of the time scale modification, the magnitude of the compensation monotonically increase according to that. On the other hand, the latter interpretation predicts that the speech compensation occurs only when the time-scale factor exceeds the boundary where the phonemic change occurs. The experimental result for these acoustic features support the former interpretation. However, for VF2 and TD, the motor adjustment in proportional to amount of the time scale perturbation is not clearly observed. This fact suggests that speech compensation is activated mainly by detecting the speed change in the altered auditory feedback.

The adaptation speed in the speech compensation was examined in terms of three acoustic features and in the disyllable listening test on the feedback sound. The results showed that the time courses of VF1, VF2 and TD reaches to the level of the magnitude of speech compensation in approximately three or five perturbed trials. A listening test on the feedback sound showed that speech error decreases in the five perturbed trials and approaches the score in the control condition. This indicates that the speech compensation to the altered feedback is almost complete. However, for the perturbed condition with the time-scale factor of 1.8, the score increases in perturbed trials but it is still lower than that in the control trials at the final perturbed trial. This means that the speech compensation is incomplete for the highest time-scale factor and some speech errors remain until the final perturbed trial.

5. Conclusions

We have examined speech compensation in response to time-scale altered auditory feedback, where the spectral dynamics are altered in the auditory feedback. Experimental results showed that speech compensation was occurred in VF1, VF2 and TD. Speech compensation is carried out mostly by changing the motion speed in the jaw opening and partially by changing the tongue motion speed and adjusting the onset time during the transition of the tongue motion to the jaw opening. Furthermore, it suggests that the speech compensation is activated mostly by detecting the temporal speed change in altered feedback speech.
6. References


