

What's in the Schwa: How Japanese L2 Learners Perceive It (or don't)

SHINICHI TOKUMA

0. Introduction

The author has observed that some Japanese university students who learn English as an L2 tend to have trouble perceiving weak syllables in words, particularly when the syllable comes in a word-initial position and is made up solely of schwa /ə/. For instance, they mistake 'amend' for 'mend', because they cannot perceive the /ə/ in 'amend'. The constraints of the Japanese phonological structure, which is often a cause of difficulty in the perception of English sounds, like the syllable final consonant in /CVC/, does not apply in this case, because the /VCV/ structure is perfectly possible in Japanese, as in /ame/ 'rain / sweets'.

Then, why can't they perceive the initial /ə/? One probable explanation is the lack of vowel reduction in Japanese: the process of neutralising vowel quality contrasts in unstressed syllables (See Lindblom (1963), Koopmans-van Beinum (1980) and van Bergem (1995) for further discussions on vowel reduction). /ə/ is considered to be the product of this process, and its reduced phonetic quality can cause a problem for Japanese learners.

However, /ə/ is not acoustically identical in all phonetic environments. Flemming & Johnson (2007) demonstrates that there are two types of /ə/ in English: one produced by centralisation and the other by contextual assimilation.¹⁾ The former has a relatively consistent vowel quality, usually

1) Also, Ladefoged & Johnson (2014) states that not all weak syllables contain /ə/ and not all vowel reduction processes produce a fully weak vowel /ə/ (p. 104).

mid-central, and appears in the word-final position, while the latter is relatively high and varies contextually in backness and lip-position, occurring in the word-internal position. Unfortunately, Flemming & Johnson did not investigate the word-initial /ə/, and therefore, it is not clear whether the word-initial /ə/ falls into one of the two types stated above or which type it belongs to if it does.

This study investigates the two issues raised above: (1) how the word-initial /ə/ in English is phonetically realised – specifically as centralisation or contextual assimilation; and (2) whether the author’s personal experience of Japanese learners’ difficulty stated above can be confirmed in the perceptual experiment, and if it can, what acoustic property of the word-initial /ə/ makes it difficult for Japanese learners to perceive /ə/.

In this paper, three experiments are described: an acoustic experiment to investigate the acoustic property of the word-initial /ə/ in English, and two perceptual experiments where Japanese university students who learn English as L2 identify the acoustically-modified word-initial /ə/.

1. Experiment 1 – Acoustic Analysis

1.1 Recording

Six natural English words which have the /əCV/ structure in the word-initial position and have a corresponding natural word with the same /CV/ were chosen for the experiment. They are: attend (tend), address (dress), appoint (point), amend (mend), account (count), agree (green). (Words in brackets are corresponding words without the word-initial /ə/.) The words were changed into past tense (e.g. address → addressed) and put in two types of frame sentences: ‘We are looking for someone who _____ it.’ and ‘I can’t believe he said he _____ it.’ This provides two vowel contexts for /ə/ for each following consonants: /u:əC/ and /i:əC/. The corresponding /CV/ words (i.e. those without the word-initial /ə/) were also put in the two types of frame sentences and they were also recorded for reference.

A male RP speaker in his forties, with no speech or hearing disorder, participated in the recording session. He was asked to read the whole frame sentence twice at a comfortable speed. The recording was made in a quiet room with Sony Electret Condenser Microphone ECM-PCV80U attached to a Windows laptop PC, and the sound was transferred to the PC by Praat (ver. 6.0.43), at a sampling frequency of 44100 Hz.

1.2 Analysis

As mentioned in Recording, Praat was used for analysis. The first and second formant frequencies were measured at the mid-point of /ə/ in each target word, as well as the intensity peak values in dB and duration of /ə/. The mean values were calculated across two repetitions, but not across each consonantal context or type of preceding vowel.

1.3 Results and Discussion

Table 1 below shows the mean formant values of the initial /ə/, separated according to the preceding vowels.

The values demonstrate clear differences between the two conditions, i.e. preceded by /u:/ and /i:/, and the mean F1 and F2 values do not correspond to those of a prototypical central vowel, 500Hz and 1500Hz respectively. Moreover, some influence of contextual assimilation can be observed: the F2

Table 1: Mean F1/F2 values of the initial /ə/ in Hz

	mean F1/F2 values (in Hz)	
	/u:əC/	/i:əC/
appoint	395/1076	428/1661
amend	487/1365	390/1879
attend	470/1669	406/1846
address	362/1661	369/1948
account	483/1503	377/2097
agree	372/1267	366/1976

Table 2: Mean intensity peak and duration of the initial /ə/

	mean intensity peak (in dB)		mean duration (in ms)	
	/u:əC/	/i:əC/	/u:əC/	/i:əC/
appoint	62.5	67.8	69.6	72.1
amend	60.5	68.8	71.7	73.7
attend	69.4	65.2	62.9	69.6
address	67.5	66.6	67.1	67.3
account	63.8	65.0	83.1	78.9
agree	71.5	67.3	66.3	72.9

values of each /ə/ are closer to those of the preceding vowels,²⁾ and when the /ə/ is followed by an alveolar consonant, the F2 values are closer to the locus value of the consonant, which is 1800Hz. This means that the initial /ə/ is not a product of centralisation and it is more likely to be a product of contextual assimilation.

The mean intensity and duration of the initial /ə/ are shown in Table 2 below.

Table 2 demonstrates that there is little difference between the mean duration values and mean peak intensity values according to the consonantal context. In contrast, as shown in Table 3 below, across the two consonantal contexts, while the mean intensity peak values are not so different from the values of the following vowels, the mean duration values show a clear difference between /ə/ and the following vowels – the former is shorter than the latter.

From this, it is hypothesised that shorter duration of the word-initial /ə/ plays a greater role in Japanese learners' problems with its perception than the other acoustic parameters. Hence, a perceptual experiment was designed

2) Typical F2 values of /u:/ and /i:/ in RP are around 900 Hz and 2300 Hz, respectively. See Hawkins & Midgley (2005).

Table 3: Mean intensity peak and duration of the initial /ə/ and the following /V/

	mean intensity peak (in dB)		mean duration (in ms)	
	/ə/	following /V/	/ə/	following /V/
appoint	65.2	70.8	72.1	122.7
amend	64.7	72.6	72.7	104.3
attend	67.3	67.5	65.6	100.9
address	67.1	72.0	67.2	91.9
account	64.4	69.3	81.0	155.4
agree	69.4	72.9	74.9	105.2

to examine whether durational change influences the perception of the word-initial /ə/. Furthermore, two groups of Japanese learners with different English proficiency levels were chosen as the participants to investigate whether their proficiency affects the perception. The details of the experiment are described in the next section.

2 Experiment 2 – Perceptual Experiment

2.1 Participants

Two groups of participants joined the experiment.

(A) Low advanced Japanese learners of English (henceforth called AJ): thirteen Japanese first-year university students of the Faculty of Commerce, Chuo University in Tokyo. All of them belonged to a special English class designed to prepare them for study abroad. Three of them had lived in an English-speaking country for at least one year. Their mean screening-test (CASEC) score before the term started was 702 (TOEIC L&R equivalent 750). This and informal evaluation by the author and the other lecturer of the class established their English abilities around the low-advanced level (CEFR B2–C1).

(B) False-beginner Japanese learners of English (henceforth called FJ): twenty-eight Japanese first-year students of the Faculty of Commerce,

Chuo University. None of them had lived or studied abroad, and their mean screening-test (CASEC) score was 247 (TOEIC L&R equivalent 205). This, and informal evaluation by the author, established their English abilities around the false-beginner level (CEFR A1).

2.2 Materials

All the pairs of words with the word-initial /ə/ were re-examined in the recording used in the previous acoustic analysis in Section 1, in order to find an optimal token for stimulus manipulation, and one of the ‘appoint’ tokens in the frame sentence ‘I can’t believe he said he _____ it’ was selected for the following reasons: 1) this token has a reasonably clear boundary between /ə/ and the preceding vowel (because /ə/ was immediately preceded by either /u:/ or /i:/ in the frame sentence, in many cases the boundary between two segments was difficult to identify due to coarticulation), and 2) it has relatively steady formant trajectories, facilitating the easier manipulation of duration.

The original duration of /ə/ in the target word is 74.6ms and it was varied in 5 steps of 20ms from 0ms to 75ms, i.e. 0-20-40-60-75. The 0ms token meant it had no word-initial /ə/, and this was to make the whole word sound like ‘point’. This is confirmed by a casual listening session with a native speaker of SSBE (Standard South British English).

The actual process of durational manipulation was as follows. First, using Praat, the target /ə/ in the frame sentence was labelled and segmented and the duration tier was created on the manipulation editor, with the onset and coda of /ə/ marked. The zero-cross points of the waveform were selected to avoid clipping when the boundaries of /ə/ were marked. Then the duration of /ə/ was manipulated by making a script to change the values of the duration tier and replacing it with the original duration tier. By doing this, the influence of the durational change on other acoustic parameters such as F0 and formant frequencies could be averted or be kept to be a minimum. The durational re-synthesis output was checked. See Figure 1 for the

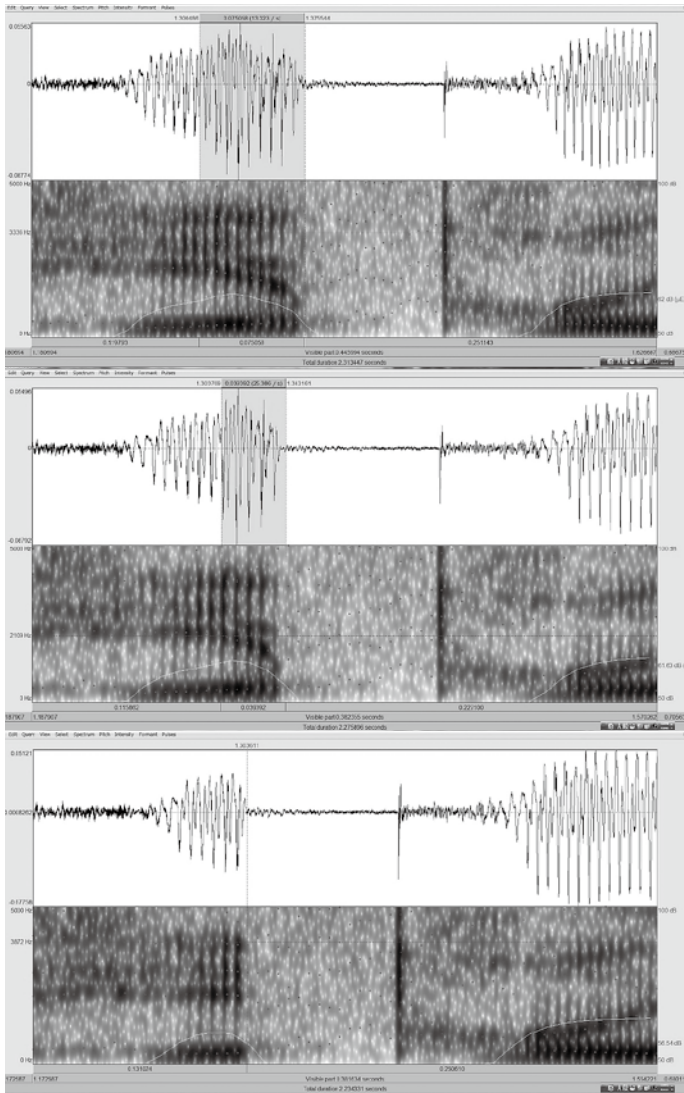


Figure 1: Spectrograms and waveforms of the onset of ‘appoint’ with /ə/ duration of 75ms (original - top), 40ms (middle) and 0ms (bottom). The parts of /sedhi:əpɔ:/ in ‘... said he appointed ...’ are shown, with highlighted parts corresponding to the edited /ə/. Note /ə/ is missing in the 0ms token.

spectrograms and waveforms of actual stimuli. The whole sentences, not the target words themselves, were presented as stimuli to the participants since the whole sentence, serving as a frame sentence to present stimuli, tells the participants the speech rate.

2.3 Procedure

The two groups of participants, AJ and FJ, were tested in a quiet Language Laboratory room at Chuo University. All the members of each group were tested at the same time, but each group was tested on a separate date. The stimuli were played through covered-ear headphones at a sound level adjusted by them, and none of them reported that their attention had been diverted by extraneous noises or by the presence of other participants.

The participants were seated individually in front of laptop PCs, and their task was to listen to the stimulus token, with different /ə/ durations, through headphones, and using a mouse, click the word in the ‘appoint’-‘point’ minimal pair that appeared on the screen as quickly as possible. The locations of the words on the screen, i.e. whether ‘appoint’ appeared on the right or left, were also randomised, and each stimulus token was played in a random order to the participants three times. In total, there were 30 presentations per participant: 5 durational patterns (0-20-40-60-75ms) x 2 button positions x 3 repetitions. The whole experiment process was controlled by Praat, utilising Experiment MFC objects.

The experiment was preceded by a task demonstration by the author designed to make the participants familiar with the experimental setting and the nature of the stimuli. Care was taken that the participants knew that they must not click a word on their screen before it was played.

2.4 Results and Discussion

After the experiment, it was found that one participant in the FJ group had not completed the task. The participant, therefore, was excluded from the data analysis, and this reduced the FJ group to twenty-seven members.

In the analysis, the number of the ‘point’ and ‘appoint’ choices was counted for each duration type across all the subjects within the group, before the percentages of the ‘point’ / ‘appoint’ choices were calculated. Figures 2 and 3 show the percentages of the choices for each durational pattern and for each subject group. Figure 2 is for the AJ group and Figure 3 for the FJ group. In these Figures, durational values of /ə/ are plotted on the X axis, and the percentage of responses on the Y axis.

Figure 2 demonstrates that the AJ group were sensitive to the duration of /ə/ and were able to identify the word-initial /ə/, which is shown as a clear categorical perception pattern, with the boundary around 20ms. This means that low-advanced learners (and probably mid- and high-advanced learners) have no difficulty in identifying the word-initial /ə/.

This finding does not conflict with what was stated in the Introduction: some Japanese L2 learners have trouble perceiving the word-initial /ə/. Although Figure 3, the results of the FJ group, shows a very weak categorical curve, which supports the role of duration in the perception of the word-initial /ə/, 37% of the stimuli were still perceived as ‘point’ even though the /ə/ had a duration of 75ms, the original value of ‘appoint’. This suggests that the initial /ə/ was difficult for the FJ group to identify and confirms the author’s personal experience.

The difficulty that the FJ group faced when perceiving the initial /ə/ is reflected in the individual finding: three of the FJ group could not perceive the /ə/ at all in all six presentations (2 button positions x 3 repetitions) of the token with full (i.e. 75ms) /ə/ duration, and two perceived /ə/ only once in six presentations. In contrast, 87% of the token with no /ə/ (i.e. 0ms) was perceived as ‘point’, with no participant perceiving the token as ‘appoint’ in all or five of the six presentations.

One might argue that the results of this experiment are not definite enough to support the statement that Japanese L2 learners have trouble perceiving the schwa, but the results of the AJ group demonstrate text-book categorical perception curves, meaning that they had no trouble identifying

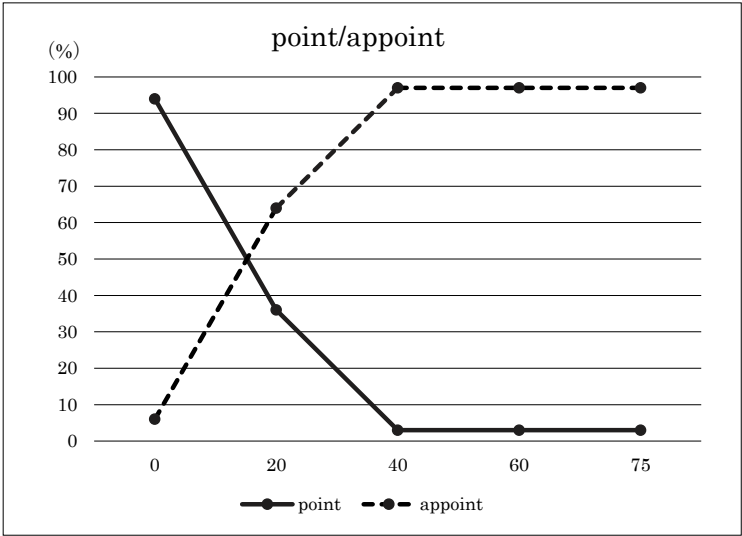


Figure 2: Results for AJ Group

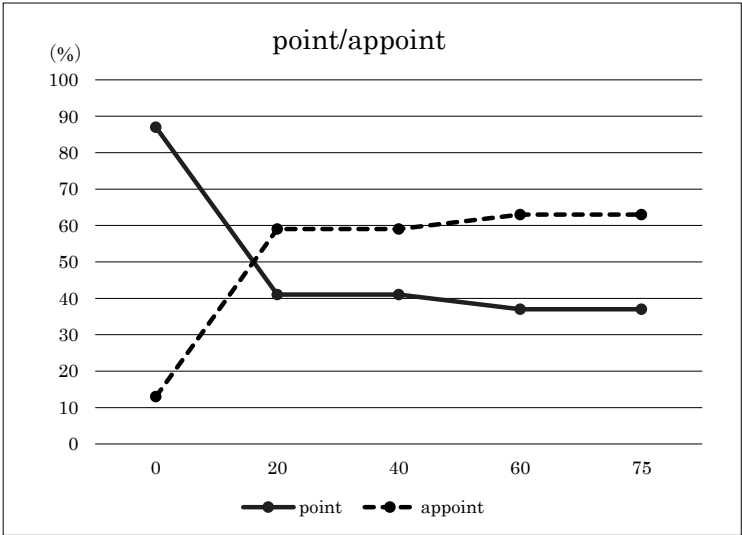


Figure 3: Results for FJ Group

the initial /ə/. Another potential criticism concerns the consonantal context – only /p/ was tested. To address these potential criticisms, a third experiment was designed testing the comprehension of ‘tend’-‘attend’, to investigate whether the results of the first experiment could be reproduced in other stimuli. Part of the manipulation process of the stimuli was changed. The details are described in the following section.

3 Experiment 3 – Perceptual Experiment

3.1 Participants

As in Experiment 2, two groups of participants, AJ and FJ, participated in the experiment: the same 13 students and 25 of the FJ group who had taken part in the first experiment. This experiment was performed on a separate date from Experiment 2. The English abilities of the FJ Group in this experiment are judged to be around CEFR A1, the same as in the first perceptual experiment.

3.2 Materials

To examine the influence of the consonantal frame, one of the ‘attend’ tokens in the frame sentence ‘We are looking for someone who _____ it’ was selected. However, in this experiment, a token with slightly dynamic formant trajectories and a clear transition from /ə/ to /u:/ was chosen for the following reason: The dynamic information available in the formant transition into and out of the quasi-steady nucleus is crucial for coarticulated vowel perception, as Strange, Jenkins & Johnson (1983) and Strange (1987) demonstrated. If the transition to the preceding /u:/ from the target /ə/ is included in the manipulation, this will shorten the transition and reduce the perceptual cue to the following /ə/, making it more difficult for listeners to identify the schwa.

The original duration (i.e. excluding the transition to the preceding /u:/) of /ə/ in the target word was 61.5ms and it was decided to manipulate the final 75ms segment of /u:ə/ continuum including the transition. The value

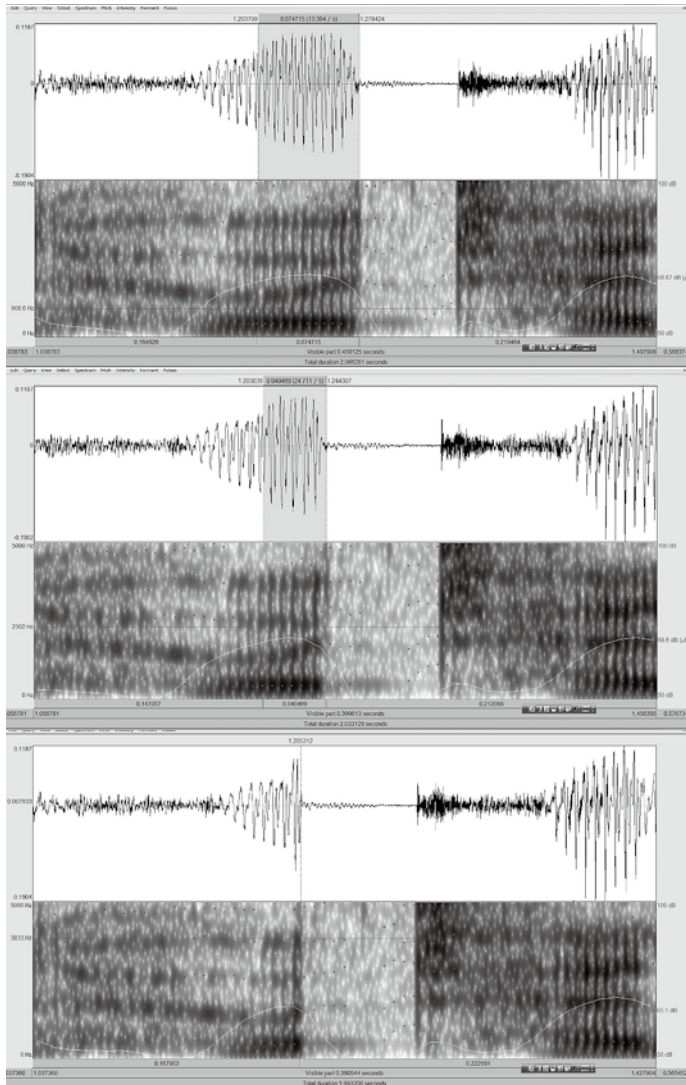


Figure 4: Spectrograms and waveforms of the onset of ‘attend’ with /ə/ duration of 75ms (top), 40ms (middle) and 0ms (bottom). The parts of /hu:ətə/ in ‘... who attended’. are shown, with highlighted parts corresponding to the edited /ə/.

of 75ms was chosen to make it in line with the previous experiment. The segment was varied in 5 steps from 0ms to 75ms, i.e. 0-20-40-60-75. As in the previous experiment, the same native speaker of SSBE (Standard Southern British English) was asked in a casual listening session to judge the 0ms token and it was perceived as 'tend'.

The actual process of durational manipulation was identical to that in Experiment 2: using Praat script, the duration of the target /ə/, including the transition from the preceding /u:/, in the frame sentence was manipulated. The results of the durational re-synthesis were also checked. See Figure 4 for the spectrograms and waveforms of actual stimuli. As in the previous experiment, the whole sentences were presented as stimuli to the participants.

3.3 Procedure

It was identical to the previous experiment: the AJ and FJ group were tested in the same Language Laboratory room at Chuo University. The participants sat individually in front of laptop PCs, and they were told to listen to the stimulus token, with different /ə/ durations, through headphones, and using a mouse, click the word in the 'tend'-'attend' minimal pair that appeared on the screen as quickly as possible. There were, as in Experiment 2, 30 presentations per participants: 5 durational patterns (0-20-40-60-75ms) x 2 button positions x 3 repetitions. The whole experiment process was controlled by Praat, utilising Experiment MFC objects.

3.4 Results and Discussion

It must be noted that all the participants in this experiment completed the task, unlike in Experiment 2, and all the data was used in the analysis.

As in Experiment 2, the responses were accumulated and the number of the 'tend' or 'attend' choices was counted for each duration type across all the subjects within the group, before the percentages of the 'tend' / 'attend' choices were calculated. Figures 5 and 6 show the percentages of the choices for each durational pattern and for each subject group. Figure 5 is for the AJ

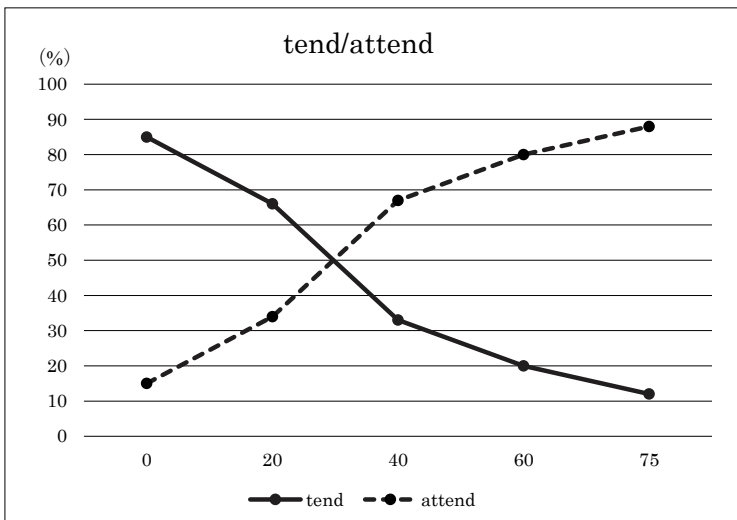


Figure 5: Results for AJ Group

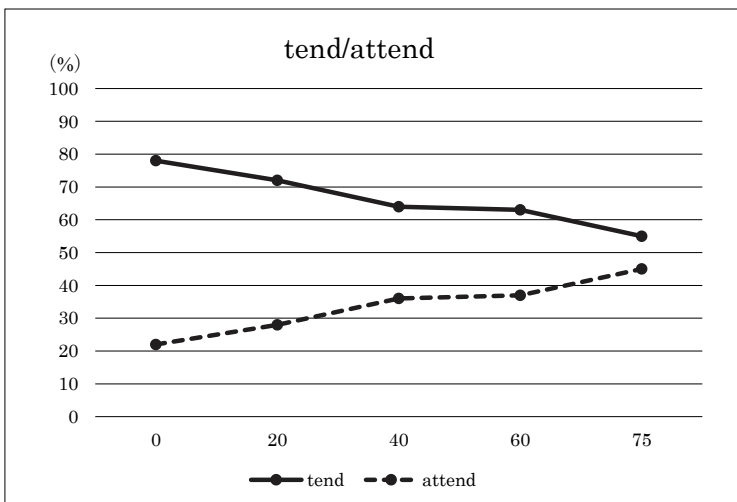


Figure 6: Results for FJ Group

group and Figure 6 for the FJ group. In these Figures, durational values of /ə/+transition are plotted on the X axis, and the percentage of responses on the Y axis.

Figure 5 demonstrates that, as in the previous experiment, the AJ group were sensitive to the duration of /ə/ and were able to identify the word-initial /ə/. However, this time, the categorical curves are much less sharp, with the boundary shifted to the point between 20ms and 40ms. This suggests that although low-advanced learners can identify the word-initial /ə/ in this context, the editing and partial deletion of the transition confuses their perception. This is in agreement with the results of Strange and others, who claim the importance of transitions in the perception of vowels. Moreover, the boundary shift can be explained by the duration of the remaining transition, $75-61.5=13.5$ ms.

The results of the FJ group in this experiment, shown in Figure 6, are in stark contrast to those obtained in the previous experiment: there is no categorical boundary, and the numbers of 'tend' choices are higher across all durational patterns, even in 75ms token with full word-initial /ə/ duration. This seems to prove that the FJ group have great difficulty in perceiving the word-initial /ə/, and the author's personal experience of Japanese learners' difficulty with this, stated in the Introduction, is confirmed. A very weak effect of duration can also be observed in Figure 6: the number of 'attend' responses increases as the duration becomes longer, suggesting that the FJ group are slightly sensitive to the /ə/ duration, but their responses are not sufficiently strong to form a categorical pattern.

There is one issue that needs to be accounted for: why the FJ group performed significantly worse in this experiment. One possible explanation is the difference in the consonantal / vowel context. Another, more plausible explanation is that since the duration of the /ə/ itself was shorter in this experiment due to the inclusion of the transition, this, together with the partial deletion of the transition, made the /ə/ identification more challenging for them because, due to their low English proficiency, they

are more easily influenced by the lack of acoustic cues. This explanation, however, must be examined further in a future experiment with more acoustically-controlled materials.

4. Conclusion

The results of the three experiments address the research questions presented in the Introduction by demonstrating that: 1) the word-initial /ə/ in English is phonetically realised as contextual assimilation, and 2) Japanese learners, particularly with low English proficiency, have difficulty in perceiving the word-initial /ə/ in English, and the durational difference contributes to the perceptual performance – the shorter the duration is, the more difficult the perception becomes. It was also found that a difference in English proficiency affects the perception – the advanced learners are more sensitive to the word-initial /ə/ itself, as well as to the transition to it. Further research is required to establish why the word-initial /ə/ in some phonetic contexts is easier to perceive for Japanese learners.

References

- van Bergem, D. (1995) *Acoustic and Lexical Vowel Reduction*. PhD Dissertation, University of Amsterdam.
- Boersma, P. & Weenink, D. (2018) *Praat: doing phonetics by computer* [Computer program]. Ver. 6.0.43. Retrieved in September 2018 from <http://www.praat.org/>.
- Flemming, E. & Johnson, S. (2007) "Rosa's roses: Reduced vowels in American English." *Journal of the International Phonetic Association*. 37: 83–96.
- Hawkins, S. & Midgley, J. (2005) "Formant frequencies of RP monophthongs in four age groups of speakers." *Journal of the International Phonetic Association*. 35: 183–198.
- Koopmans-van Beinum, F. J. (1980) *Vowel contrast reduction: an acoustic and perceptual study of Dutch vowels in various speech conditions*. PhD Dissertation, University of Amsterdam.
- Ladefoged, P. & Johnson, K. (2014) *A Course in Phonetics*. 7th ed. Cengage Learning.
- Lindblom, B. (1963) "Spectrographic Study of Vowel Reduction." *Journal of the Acoustical Society of America*. 35: 1773–1781.
- Strange, W. (1987) "Information for vowels in formant transitions." *Journal of Memory and Language*. 26: 550–557.
- Strange, W., Jenkins, J.J. & Johnson, T. L. (1983) "Dynamic specification of coarticulated vowels." *Journal of the Acoustical Society of America*. 74: 695–705.