An empirical study on the effects of acoustic and semantic contexts on perceptual learning of L2 phonemes

Yuko IKUMA
Graduate School, Kobe University / ATR Human Information Science Lab

Reiko AKAHANE-YAMADA
ATR Human Information Science Lab / Graduate School, Kobe University

Abstract

A laboratory training study was conducted in order to investigate the effect of context on identification of English words minimally contrasting in /tr/-/ʌl/, /b/-/v/, or /s/-/θ/. Thirty native speakers of Japanese participated in a word identification task and perceptual training under the three stimulus conditions: (1) WD condition (words in isolation), (2) NS condition (words within a semantically neutral carrier sentence), and (3) CS condition (words within a semantically contextual carrier sentence). Identification performance for /tr/-/ʌl/ and /b/-/v/ during test phases showed that the accuracy differed significantly between the stimulus condition in the order, NS < WD < CS. This tendency was not significant for the /s/-/θ/ contrast, which showed the highest identification score among the three contrasts. These results suggest that semantic context raised identification accuracy while acoustic context lowered it when the target phonemes were difficult to identify. Furthermore, in case of /tr/-/ʌl/, it was shown that training with the CS stimuli improved the accuracy for stimuli in CS condition but not for those in other conditions. This result together with the above results implies that when language learners are provided with meaningful information during training, they are likely to learn the strategy to exploit semantic cues. Thus, perception training based on auditory input is revealed to be important in order to develop robust listening comprehension ability.

1. Introduction

Both phonological decoding and semantic decoding are involved when a spoken language is processed. Many studies have reported that native listeners' phoneme identification accuracy is affected by semantic contexts (e.g., Morton & Long, 1976; Craig, 1988; Burton, Baum, &
Blumstein, 1989; Griffin & Bock, 1998). This effect was further examined for elderly subjects (Dubno, Ahlstrom, & Horwitz, 1999). Furthermore, Mayo, Florentine, & Buus (1997) showed that when listening speech under noise, native-monolingual listeners and early bilingual listeners utilized contextual information much more than late bilingual listeners did. However, not many studies have examined the effect of contextual information for L2 (second language) listeners.

Thus, Rothwell & Akahane-Yamada (2003) examined the effect of acoustic and semantic contexts on Japanese listeners' identification of English words. English words that minimally contrasted in /r/ and /l/ (e.g., right vs. light) were produced by native speakers of American English in isolation, within semantically contextual carrier sentences, and within semantically neutral carrier sentences. Result showed that identification accuracy of words decreased in semantically neutral sentences but improved in semantically contextual sentences, suggesting that phoneme recognition is hindered by acoustic context, but is facilitated by semantic context.

The present paper replicated and expanded Rothwell & Akahane-Yamada (2003) in two ways. First, we tested words contrasting in /r/-/l/ and /s/-/θl/, as well as /r/-/l/. Second, we evaluated the effectiveness of training stimuli in the three contextual conditions, WD (in isolation), CS (within contextual sentence), and NS (within neutral sentence), through a laboratory training experiment.

2. Method

2-1. Stimuli

The stimuli were 45 pairs of English words minimally contrasting in /r/-/l/ (RL), 45 pairs contrasting in /b/-/v/ (BV), and 37 pairs contrasting in /s/-/θl/ (STH). Each item was used as isolated word stimuli (WD); and also as sentence stimuli, in which the item was embedded within a semantically neutral carrier sentence (NS) and within a semantically contextual carrier sentence (CS). Examples of NS and CS are shown in Table 1. All words and sentences were produced

<table>
<thead>
<tr>
<th>NS*</th>
<th>Say correct on Tuesday.</th>
<th>CS</th>
<th>I came over to correct your mistakes.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Say collect on Tuesday.</td>
<td></td>
<td>I came over to collect your garbage.</td>
</tr>
</tbody>
</table>

* The carrier sentences of NS differently assigned for every word and speaker.

| Table 2. List of the number of experimental materials in each condition |
|-----------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                             | RL (/r/-/l/)    | BV (/b/-/v/)    | STH (/s/-/θl/)  |
| Phase                      | test training   | test training   | test training   | test training   |
| Target stimuli             | 10 (5 pairs)    | 10 (5 pairs)    | 10 (5 pairs)    | 10 (5 pairs)    |
|                            | 80 (40 pairs)   | 80 (40 pairs)   | 80 (40 pairs)   | 64 (32 pairs)   |
| Sum                        | 90 (45 pairs)   | 90 (45 pairs)   | 74 (37 pairs)   |
by 5 native speakers of North American English (3 males, 2 females) and read at a self-selected normal speaking rate. The recording took place in an anechoic chamber and was digitized at 16-bit resolution and 44.1 kHz sampling frequency. Thirty items in the WD condition and their corresponding items in the NS and CS conditions (10 stimuli for each contrasting phoneme) were for test session (speaker = 1 male); the other stimuli (224 items each in the three conditions: 40 in RL, BV and 32 in STH) (see Table 2.) were divided into 4 groups and assigned to 4 training sessions (speakers = 2 males, 2 females).

2-2. Participants
The participants were thirty native Japanese speakers who had never lived outside Japan for more than three months and who were not pursuing a degree in English (22 males and 8 females, ranging from 18 to 26 years old). All of them had normal hearing ability and were paid for their participation.

2-3. Procedure
The experiments were conducted over two weeks for each participant and all the participants took a two-hour session per week. They were tested and trained with the /b/-/v/ contrast on the first day of the session. After a week, on the second day, they were exposed to the /s/-/θ/ contrast. Finally, after another week, on the last day of the session, the /r/-/l/ contrast was the target stimuli.

For each phoneme the experiment consisted of the same structure: pretest, training blocks, and post-test. The pretest operated on the same sequences as the post-test. Ten words (5 pairs) were assigned as test stimuli for each contrast. Test sessions were carried out in a fixed order as follows;

(1) Learning Vocabulary (LV): In order to control the knowledge of the target word meaning, participants took part in a 2-alternative forced-choice translation task: 3 blocks from Japanese to English (forward: FW), in which participants were to read the Japanese word shown on the computer monitor and identify the English meaning between the 2 choices, and 3 blocks from English to Japanese (backward: BW), and these two kinds of task repeated alternately. In these blocks, the stimulus words were the same as those used in the test sessions afterward in the same section, and participants received feedback and correction trials (if mis-translated) on each trial.

(2) Listening Words (LW): Words were presented aurally. The task was a 2 alternative forced-choice identification task with no feedback and no correction trials. Two words comprising a minimal pair were presented visually on the computer screen, and one of the two words was presented aurally. Participants were to select which of the two items they heard.

(3) Listening Sentences (LS): The target words within a sentence were presented aurally. Some
information concerning the target sentence was displayed on the screen visually. A series of rectangles were presented visually; each rectangle represented one word in the sentence, and a red rectangle was used to indicate the position in which the target words appeared. The task was a 2AFC task without any feedback or correction trials. Each word was presented twice on different trials: once within a neutral carrier sentence, once within a contextual carrier sentence.

(4) Visual Sentence (V-Sent): In order to evaluate the contextuality of the sentences used in the test, a supplementary task was carried out. A carrier sentence was presented on the screen with the target word blanked out. No auditory stimuli were presented. Each word appeared twice; once within a neutral carrier sentence, and once within a contextual carrier sentence. The task was a 3-alternative forced-choice task, in which the choices were 2 words of a minimal pair and “either” for neutral carrier sentences.

For the training session, 80 RL and BV words (40 pairs), and 64 STH words (32 pairs) were assigned and they were divided equally into 4 sub-sessions. Training sessions consisted of two tasks:

(1) Learning Vocabulary (LV): This was the same as the LV task in the test session.
(2) Listening Training (LW or LS): Participants received 3 kinds of perceptual training: training of listening to words in isolation (WD training), listening to neutral carrier sentences (NS training), and listening to contextual carrier sentences (CS training). Only one kind of training was used for each phoneme contrast (RL, BV and STH), and the order of training was counterbalanced among participants.

3. Results

Note first that all participants reached 100 % in accuracy during 6 blocks of LV (translation task) conducted prior to other tasks. This assures that at least the target words within the sentences activated correct semantic encoding in the later tasks.

Identification accuracy averaged over the stimuli was computed for each phoneme contrast, for each pre- and post-test and for each stimulus condition. Separate ANOVA was conducted for three phoneme contrasts, where test phase (pretest and post-test), test stimulus condition (WD, NS and CS), and training stimulus condition (WD, NS and CS) were the variables.

3-1. RL Identification

Concerning perception of the RL contrast, the main effect of test phase, the main effect of test stimulus condition, and the interaction among three variables (test phase, test stimulus condition and training stimulus condition) were statistically significant. Together with the result of multiple comparison tests on the main effects, it was shown that the identification accuracy
improved from the pretest to the post-test and significantly differed among the stimulus conditions in the following order; NS < WD < CS (Figure 1; left).

**The effect of training**

Considering the interaction among all the factors, a post-hoc test was conducted. The result of the post-hoc test showed that only the accuracy for CS stimuli by the CS training group improved significantly (Figure 2). No significant effect was observed for the other test stimulus condition nor for the other training groups; except for WD stimuli and NS stimuli of the NS training group, for which only a mild improvement was seen.

**3-2. BV Identification**

Analysis of variance on BV contrast showed that the main effect of test stimulus condition and the interaction between test stimulus condition and test phase were significant. The identification accuracy differed among the stimulus conditions in the following order; NS < WD
< CS, from the result of multiple comparison tests on the main effect of the factor of the test stimulus condition (Figure 1; center). Analysis of the interaction showed that accuracy differed significantly only between NS and the other two conditions (NS < WD = CS) before training, but after training, the accuracy between WD and CS arose to be significant (NS < WD < CS).

The effect of training

The main effect of any variables and their interactions were not significant.

3-3. STH Identification

Only the main effects of test phase and test stimulus condition were statistically significant for the STH contrast. Multiple comparison tests showed that the identification accuracy significantly improved from the pretest to the post-test, and that it differed significantly among stimulus conditions in the following fashion: NS < WD and CS < WD (Figure 1; right panel).

The effect of training

Again, the main effect of any variables and their interactions were not significant.

4. Discussion

Identification results during test phases showed that the acoustic context decreases but the semantic context increases the perception accuracy of /l/-\&/ by Japanese EFL (English as a foreign language) learners. The same tendency was observed for the /b/-\&/ contrast, replicating the study by Rothwell & Akahane-Yamada (2003). However, the semantic context did not improve the identification accuracy of /s/-\&/ . Note that the identification accuracy of /s/-\&/ was much higher than that of /l/-\&/ and /b/-\&/. We hypothesize that listeners depend on semantic information rather than acoustic information when the target word or phoneme is difficult to identify from the auditory input. If this is the case, the present phenomenon is consistent with the results of previous studies which demonstrated that native listeners' speech perception is affected by semantic context under noise (e.g., Morton & Long, 1976; Craig, 1988 for English speaking young adults; Dubno, Ahlstrom, & Horwitz, 1999 for young and aged adult English speaker; Fallon, Trehub, & Schnider, 2002 for children and adults; Mayo, Florentine, & Buus, 1997 for bilinguals and monolinguals of American-English, etc.).

The effect of the training stimulus condition was evaluated by the improvement of identification accuracy from pretest to post-test. Unfortunately, there was no substantial improvement, probably because of an insufficient number of training trials during the training phase. In the present study, 480 (/l/-\&/ and /b/-\&/) or 512 (/s/-\&/) trials were given for the training of each contrast, whereas one previous training study gave 4,352 trials to attain an
improvement of 8 points in accuracy for /r/-/l/ identification (Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994), and 12,240 trials to attain an improvement of 20 points (Akahane-Yamada, 1996). Interestingly, however, there was a significant improvement for only one case. When listeners were trained to identify /r/-/l/ words using words within contextual sentences, their identification accuracy for /r/-/l/ words within contextual sentences significantly improved from pretest to post-test, but not for the words in isolation or words within neutral sentences. This result suggests that when Japanese EFL learners are trained only using materials with semantically meaningful context, the training does not guarantee perceptual improvement in the isolated-word condition or the neutral sentence condition. In fact, after the experiment, some of the participants reported that when the target words were difficult to perceive, other portions of the sentences were useful as a clue to identify the target words.

When the above two results are taken into account, we may conclude that acoustic context interferes with the identification of words in sentences, but semantic context helps when the words are difficult to identify from acoustic information alone. However, when FL (foreign language) learners are trained to identify phonemes or words by using speech that has phonemes/words within a semantically meaningful sentence, their listening ability of phonemes or words from acoustic signal may not improve. Although further investigation is necessary, the present results can make a theoretical contribution to the development of effective FL training materials.

5. Conclusion and Implications

The present study has shown with empirical evidence that acoustic context interferes with perception accuracy while semantic context facilitates phoneme identification by Japanese EFL learners, especially on difficult-to-identify phonemes. The advantage of semantic context is weakened on much easier pairs. Perceptual learning of phonemes does not progress in a short period of time, but advantage gained by semantic context appears relatively promptly and to a substantial degree.

Based on the results of the study, the importance of acoustically based training should be considered; training of the aural ability for individual phonemes is effective in cultivation of aural comprehension. Further research will be needed to examine whether materials with low semantic information is efficient for second language aural education.

Acknowledgements

We are grateful to the members of Spoken Language Acquisition Project at ATR Human Information Science Laboratories for useful discussion, data collection, data processing, and programming assistance. This research was conducted as part of Research on Human
Communication' with funding from the Telecommunications Advancement Organization of Japan.

References


