Effects of acoustic and semantic contexts in learning L2 phoneme perception: Do they help or interrupt?

Yuko Ikuma\(^{(1),(2)}\), Reiko Akahane-Yamada\(^{(2),(1)}\)

(1) Graduate School of Cultural Studies and Human Science, Kobe Univ.
(2) ATR Human Information Science Lab.
yikuma@atr.jp, yamada@atr.jp

Abstract

Perception or more widely cognition of the words and sentences of a second language (L2) requires the process of both phonological decoding and lexical decoding. Accordingly, both acoustic context and semantic and syntactic context play an important role in L2 speech perception as well as in native speech perception. In this paper, the effect of acoustic and semantic contexts in learning L2 phoneme perception was examined. The result of a laboratory training study showed that identification accuracy significantly differed among the stimulus conditions. It was also demonstrated that acoustic context decreases but the semantic context increases identification training and that meaning-focused training progresses accuracy of the perception learning on phoneme identification only in the condition in which context has semantic meaning. These results suggest that training methods that develop the ability to identify phonemes only with auditory input may be effective in cultivation of aural comprehension.

1. Introduction

Perception of spoken language requires the process of both phonological decoding and lexical decoding. Many studies have reported that semantic information compensates for the degradation or loss in acoustic information. In case of speech perception in the first language (L1), Morton and Long (1976 [1]) reported that the reaction time for identification of words under noise was affected by both semantic contexts and syntactic contexts. They hypothesize that a suitable context for the target word prompted identification and shortened the reaction time. Craig (1988 [2]) also showed that identification accuracy for words under noise was higher when the target words appeared in highly predictable contexts. This effect was further confirmed to be true for young subjects [3] as well as elderly subjects [4]. Furthermore, it was shown that when listening to English speech, English monolingual speakers and early English-Spanish bilingual speakers utilized contextual information more than late English-Spanish speakers [5]. These findings suggest that semantically predictable context facilitates understanding and processing speed on L1 sentence perception.

Recently, Rothwell and Akahane-Yamada (2003 [6]) examined the effect of contexts in second-language (L2) speech perception. Native speakers of Japanese identified English words contrasting in /r/ and /l/ (e.g. right vs. light). Results showed that the identification accuracy was lower when the target words were presented in semantically neutral sentences (NS) than when they were presented in isolation (WD). However, accuracy was highest when the target words were presented in semantically contextual sentences (CS). These results suggest that both acoustic context and semantic context play an important role in L2 speech perception. Specifically, processing load increased when acoustic context was added, while perception accuracy was promoted when semantic context was further added.

In order to further examine the effect of contexts in L2 speech perception, the present study expanded Rothwell and Akahane-Yamada [6] in two ways, (1) examination of the effect of context when identifying /b/-/v/ and /s/-/θ/ as well as /r/-/l/. (2) examination of the effectiveness of the three contextual conditions when used in training (WD, NS and CS).

2. Method

2.1. Stimuli

CONDITION: Five native speakers of American English (three males, two females) produced 45 pairs of English words minimally contrasting in /r/ and /l/ (RL), /b/ and /v/ (BV), and 37 pairs in /s/ and /θ/ (STH) in three conditions: (1) in isolation (WD), (2) within a semantically neutral carrier sentence (NS), and (3) within a semantically contextual carrier sentence (CS) (See Table 1).

SENTENCE MATERIAL: For NS sentences, 52 carrier sentences were prepared with the target word blanked out, and assigned randomly to each minimal pair of words. The assignment differed between talkers, therefore none of the NS materials were produced by multiple talkers. The position of the target words was neither the first nor the last word in the sentence. For CS sentences, the carrier sentences were unique and specific to each word. They were not interchangeable.
with each other within each pair of words. The target words appeared neither in the first nor the last position in the sentence. The numbers of words in the sentence and the position of the target word in the sentence were similar between the two sentences of each word pair. Five pairs (10 words) were selected as test stimuli for each contrast, and the rest of the pairs (80 words for /θ/-/θ/ and /b/-/v/, and 64 words for /s/-/θ/) were used for training stimuli (Table 2).

SOUND: For the tests, the 30 words were read by one American English speaker (male) in WD, NS and CS conditions at a self-selected normal speaking rate. For training, each of the 224 words was read by four speakers (two males, two females). Speakers did not overlap between test and training. These productions were recorded in an anechoic chamber and digitized at 16-bit resolution and 44.1 kHz sampling frequency.

Table 1. Stimulus Conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>R/L</th>
<th>Examples of Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>WD</td>
<td>R/L</td>
<td></td>
</tr>
<tr>
<td>NS (*)</td>
<td>R</td>
<td>What letter does arrive start with?</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>What letter does alive start with?</td>
</tr>
<tr>
<td>CS</td>
<td>R</td>
<td>When did he arrive yesterday?</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>Was he still alive yesterday?</td>
</tr>
</tbody>
</table>

(*NS) NS carrier sentence differed by word pair and talker.

Table 2. Number of words used in the experiment

<table>
<thead>
<tr>
<th></th>
<th>test phase</th>
<th>training phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>RL</td>
<td>10 words (5 pairs)</td>
<td>80 words (40 pairs)</td>
</tr>
<tr>
<td>BV</td>
<td>10 words (5 pairs)</td>
<td>80 words (40 pairs)</td>
</tr>
<tr>
<td>STH</td>
<td>10 words (5 pairs)</td>
<td>64 words (32 pairs)</td>
</tr>
</tbody>
</table>

2.2. Participants

Participants were 30 native speakers of Japanese who had never lived outside Japan for more than three months (22 males and 8 females, ranging in age from 18 to 26 years). They were undergraduate students, but none were pursuing a degree in English. All of them had normal hearing ability and were paid for their participation. The participants were divided equally into six groups and received different training scenarios.

2.3. Procedure

There were three sessions each corresponding to one of three conditions. The three sessions were conducted on three separate days with a one-week interval between days. In the first session, the /b/-/v/ stimuli were presented, and in the second and third sessions, the /s/-/θ/ and /r/-/l/ stimuli were presented, respectively.

A pretest and post-test paradigm was administered for each phonetic contrast. During the pretest and post-test, the following five tasks were administered to the participants in a fixed order.

1. Word translation training: First, word translation training was carried out, in order to control the semantic knowledge of the words used in the following tests. It was a two-alternative forced-choice (2AFC) translation task (See Komaki and Akahane-Yamada 2004 [7], in this proceeding). In the first block, the English words appeared on the computer screen and the participants identified the English meaning between two choices (Japanese-to-English: JE) and in the next block, the English words were presented and the participants identified the Japanese expression between two choices (English-to-Japanese: EJ). Each of the ten words used in the following tests appeared twice in one block, resulting in twenty trials in each block. The set of these two blocks (JE and EJ) were repeated three times; that is, participants performed six blocks in total. Participants received feedback (chime for correct answer and buzzer for incorrect answer) and performed correction trials (if mistranslated).

2. Perception of words in isolation: Identification ability of isolated words was measured with a 2AFC task. Two words comprising a minimal pair were presented visually on the computer screen, and simultaneously one of the two words was presented aurally. The participants were to select which of the two items they heard. Feedback was not provided for the responses. Ten stimuli were presented two times each, in a random order, resulting in 20 trials in total.

3. Perception of words in sentences: Identification ability of the words within a sentence was measured in a 2AFC task. Each word was embedded within two kinds of carrier sentences, NS and CS. On the screen, a series of rectangles was presented to indicate the number of words comprising the target sentence, with a red rectangle used to represent the position in which the target word appeared. There were no feedback and no correction trials. There were 20 stimuli (10 NS and 10 CS) in this block. NS and CS stimuli appeared randomly in this block.

4. Contextuality judgment: The contextuality of the test sentences was evaluated. Three response buttons were presented on a computer screen. On two buttons, two words from a minimal pair were presented in English orthography. On the remaining button, the word “either” was presented. Simultaneously, a sentence was presented in English orthography, with the target word blanked out. Participants were to choose the most suitable item from the three choices. The “either” button was expected to be selected for NS sentences, since words in each minimal pair are equally probable in a neutral context. In contrast, for CS sentences,
whichever member of the word pair that was semantically consistent with the context, was expected to be selected. In this block, there were no auditory presentations. Feedback and correctional trials were not provided for the responses. The appeared order also randomized between five kinds of NSs and 10 CSs. This block had 15 trials.

(5) Word translation test: A test was administered to confirm whether the participants activated their semantic knowledge of the words during the perception test and contextuality judgment blocks. There were one JE block and one EJ block, and each block had no feedback and no correctional trials.

For the training session, 80 RL and BV words (40 pairs each) and 64 STH words (32 pairs) were used. For each phoneme training session, the stimuli were divided equally into four sub-sessions. Training sessions consisted of the following two tasks:

(1) Word translation training: The same task used in the test session was used.
(2) Perception training: A 2AFC identification task was used. On each trial, the two members of a minimal pair were displayed in English orthography on the CRT monitor. Simultaneously, a test stimulus was played over headphones. Participants identified the target word and were to choose the corresponding word from the two alternatives by clicking one of the buttons. Feedback was provided immediately after the response: chime for correct answer, and buzzer for incorrect answer. There were correction trials if incorrect responses were made. Twenty words in isolation (WD condition), in neutral sentence (NS condition) or in contextual sentence (CS condition) were presented in a random order in one RL training block and one BV block, and 16 words in one STH block. Four native speakers of English read different 20 (for RL and BV) and 16 (for STH) words and sentences. There were 12 RL and BV blocks and 16 STH blocks, resulting in 480 trials for RL and BV, and 512 trials for STH in total. Only one kind of training was used in each session (RL, BV and STH), and the order of training (WD, NS or CS) was counterbalanced among participants.

3. Results

First of all, all the participants reached 100% accuracy during the 6 blocks of the word translation training, suggesting that the semantic knowledge of testing and training words was controlled.

Correct response rate for each stimulus was pooled over participants. Then, separate ANOVAs were conducted for the three phoneme contrasts, where test phase (pretest and post-test), test stimulus condition (WD, NS and CS), and training stimulus condition (WD training, NS training and CS training) were the variables.

3.1. RL contrast

The main effect of the test phase [F (1, 27) = 11.7, p < 0.005], the main effect of test stimulus condition [F (2,
the interaction among all the three variables (test phase, test stimulus condition and training stimulus condition) \[ F(4, 54) = 3.02, p < 0.05 \] were significant. The result of the multiple comparison tests by Ryan’s method on the main effects showed that the identification accuracy improved from the pretest to the post-test and differed among the stimulus conditions in the following order; NS < WD < CS (MSe = 0.0184, p = 0.05, See Figure 1; left).

A post-hoc test showed that only the accuracy for CS stimuli in the CS training group improved significantly (MSe = 0.0199, p = 0.05, See Figure 2). No significant effect was observed for the other training groups or the other test stimulus conditions in the CS training group.

3.2. BV contrast

The main effect of test stimulus condition \[ F(2, 54) = 14.7, p < 0.001 \] and the interaction between test stimulus condition and test phase \[ F(2, 54) = 3.22, p < 0.05 \] were significant (Figure 1: center). Analysis of the interaction showed that accuracy differed significantly only between NS and the other two conditions before training (NS < WD = CS), but after training, the accuracy differed among all three conditions (NS < WD < CS).

3.3. STH contrast

The main effect of the test phase \[ F(1, 27) = 6.56, p < 0.05 \] and the main effect of test stimulus condition \[ F(2, 54) = 6.43, p < 0.005 \] were significant. Multiple comparison tests showed that the identification accuracy significantly differed in the following fashion: NS < WD and CS < WD (MSe = 0.0141, p = 0.05, Figure 1; right).

4. Discussion

The present result replicated the previous study ([6]): when Japanese young adults identify English /r/-/l/, the acoustic context decreases but the semantic context increases the perception accuracy. The same tendency was shown for the /b/-/v/ contrast. However, for the /s/-/θ/ contrast, a semantically contextual context did not improve the accuracy. Considering that the identification accuracy of the /s/-/θ/ words used in this study was higher than that of the other two contrasts, we hypothesize that listeners depend on semantic clues rather than acoustic features when the target phoneme was perceptually difficult to identify. On the other hand, when the target phoneme was rather easy to identify, listeners effectively utilize acoustic feature.

Regarding the effect of perception training, we could not obtain a substantial overall training effect possibly because of the lack of sufficient training trials. Interestingly, however, when Japanese speakers were trained on /r/-/θ/ contrast using semantically contextual sentences, identification accuracy improved from pretest to post-test only for contextual. This result suggests that when the learners are trained to identify phonemes using stimuli with semantically meaningful context, they may activate a strategy to exploit semantic information. As a result, they tend to pay less attention to the auditory input, and accordingly, perception learning of phonemes is interrupted.

5. Conclusion

The present study demonstrated that semantic context helps L2 listeners to identify difficult phonemes. However, there is a possibility that training with semantically contextual sentences may decrease the training effect for phoneme perception. This suggests that acoustically based training is important at least in some stage during L2 listening training.

6. Acknowledgements

This research was conducted as part of ‘Research on Human Communication’ with funding from the Telecommunications Advancement Organization of Japan.

7. References