

Title	Dependence on age of interference with phoneme perception by first- and second-language speech maskers
Author(s)	Kubo, Rieko; Akagi, Masato; Akahane-Yamada, Reiko
Citation	Acoustical Science and Technology, 36(5): 397-407
Issue Date	2015-09-01
Type	Journal Article
Text version	author
URL	<a href="http://hdl.handle.net/10119/12997">http://hdl.handle.net/10119/12997</a>
Rights	Copyright (C)2015 The Acoustical Society of Japan, Rieko Kubo, Masato Akagi, Reiko Akahane-Yamada, Acoustical Science and Technology, 36(5), 2015, 397-407. <a href="http://dx.doi.org/10.1250/ast.36.397">http://dx.doi.org/10.1250/ast.36.397</a>
Description	



# Dependence on age of interference with phoneme perception by first- and second-language speech maskers

Rieko Kubo<sup>1,\*</sup>, Masato Akagi<sup>1,†</sup> and Reiko Akahane-Yamada<sup>2,‡</sup>

<sup>1</sup>*School of Information Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan*

<sup>2</sup>*ATR Intelligent Robotics and Communication Laboratories, 2-2-2 Hikaridai, Seika-cho, Sorakugun, Kyoto 619-0288, Japan*

(Received ??? 00 0000, Accepted for publication September 99 9999)

## Abstract:

This study investigated the differences in first-language-based (L1-based) phonetic processing for second language (L2) phonemes among different age groups of adults. A speech-in-speech masking paradigm was utilized to examine the contribution of the L1-based processing. A phoneme identification task in one language was conducted in the presence or absence of an interferer of a masker of the same or a different language. The degree of interference (i.e., the decrease in identification performance) was postulated to increase as the similarity of underlying processes for the target and masker increases. Experiment 1 was conducted to test the effectiveness of the paradigm. As expected, the interference increased as the similarity of underlying processes for the target and masker increased. Experiment 2 examined the perception of English /r-/l/ and other phonetic contrasts by Japanese listeners in various adult age groups, to examine whether the degree of interference differs depending on the putative degrees of L1-based processing and on age. The results demonstrated such differences and showed that the L1-based processing can be estimated from the decrease in the identification performance. They also suggested that the perception of /r-/l/ in the initial singleton and initial cluster positions was high L1-based in older adults.

**Keywords:** Speech perception, Second language, Phonetic processing, Aging effect, Influence of first language, Speech-in-speech masking, Informational masking

**PACS number:** 43.71.Hw, 43.71.Es, 43.66.Dc, 43.71.Lz,

## 1. Introduction

Adults have been widely reported to be less successful at learning second language (L2) phonetic contrasts than children. One common explanation for the adult-child difference has been that long-lasting exposure to their first language (L1) alters speech processing and that L2 language experience can alter the listener's processing of phonemes in childhood, however, these effects have little influence in adulthood [1]. Studies of perceptual training to identify L2 phonetic contrasts [2-4] provided results that led to new insights into this idea. The results demonstrated that perceptual training improved the performance of young adults (those who were college-age or in their 20s) in identifying L2 phonemes. Subsequent studies have further demonstrated that the perceptual training improved the performance of older adults (from those in their 30s to those in their 60s) as

well [5, 6]. However, the improvements were not equal among the age groups. Studies of perceptual training that involved Japanese adults in age groups from the 30s to the 60s to identify American English /r-/l/ demonstrated that the training effectiveness gradually decreased in the older age groups [6]. The analysis of the training effect revealed that the older age groups were more strongly influenced by the positions of /r-/l/ in the word than were the younger age groups. The positions of /r-/l/ (phonetic environments) were final singleton (FP), final cluster (FC), intervocalic (IN), initial singleton (IP), and initial cluster (IC) positions (e.g., FP: "bare-bail," FC: "cold-cord," IN: "oreo-oleo," IP: "rate-late," IC: "crock-clock"). Younger age groups improved in all phonetic environments. However, the oldest age group improved for FP but did not improve for IP or IC. The improvements for FP were similar among age groups but improvements for IP and IC decreased with age.

Regarding the influence of a phonetic environment, an attempt to apply the listener's L1 phonetic categories

\* e-mail: rkubo@jasit.ac.jp

† e-mail: akagi@jaist.ac.jp

‡ e-mail: yamada@atr-lt.jp

§ An earlier version of this study was presented at the ASJ Spring Meeting, 2014

has been suggested to result in difficulties to identify /r-/l/ in IP and IC for native speakers of Japanese. A group of Japanese speakers identify FP more easily than IP and IC [2, 3, 7]. In addition, one study [8] reported that the phonetic environment's influences differ depending on the listener's L1 (Japanese vs. Korean), on the assimilation patterns to the listener's associated L1 phonetic categories, and on identification difficulties. Where the listeners assimilate an L2 phonetic contrast with two different L1 phonetic categories, the difficulties were small. Where the listeners assimilate an L2 phonetic contrast into one identical L1 phonetic category, the difficulties were large. For native speakers of Japanese, phonetic contrasts in FP and FC were assimilated into two different Japanese categories, and IP and IC were assimilated into one Japanese category. Native speakers of Korean showed different assimilation patterns and difficulties regarding the phonetic environment.

The listener's language-specific perceptual pattern likely plays an important role in the differences in the effectiveness of the training. The older adults in the perceptual training did not improve in a phonetic environment where native speakers of Japanese have difficulties identifying (IP and IC) [6]. Thus, one possible explanation for the age-related decrease in the training effectiveness is that phonetic processing for L2 phonemes is gradually altered to be high L1-based in adulthood, which brings about an age-related decrease in the training effectiveness.

The purpose of this study was to investigate the differences in L1-based phonetic processing for L2 phonemes across adult age groups before perceptual training. If the degree of L1-based phonetic processing differed among various age groups before training, it can be the starting point to examine the explanation of the differences in the training effect in future research. The degree of contribution of the L1-based phonetic processing was estimated with adults of several age groups to examine the differences.

The strategy adopted to estimate the degree of the contribution of the L1-based phonetic processing was to interfere with the access to phonetic encoding, where the processing emerges. The interference will result in a decrease in identification performance. The decrease should reflect the degree of the contribution of the L1-based phonetic processing. A speech-in-speech masking paradigm was utilized to interfere with the phonetic encoding. A speech-in-speech masking paradigm is known to produce informational masking in addition to energetic masking [9–11]. The degree of interference (i.e., the decrease in identification performance) was postulated

to increase as the similarity of the underlying processes for the target and masker increases.

Language-specific information of a speech masker would interfere with L1-based phonetic processing. Previous studies of speech-in-speech perception [9–11] indicated that target-masker (acoustic) similarity and the consistency of the language of the masker with the listener's L1 should influence decreases in identification performance. Although these studies mostly dealt with sentence perception, on the basis of the studies, the target-masker (acoustic) similarity can be assumed to interfere with the process of perceiving the acoustic features, and speech in the listener's L1 can also be assumed to interfere with the process of L1-based phonetic encoding in phoneme perception.

When the target is an L2 phoneme, L2 speech possesses high target-masker similarity but little consistency with the listener's L1. In contrast, L1 speech possesses low target-masker similarity with L2 phonemes but high consistency with the listener's L1. Thus, a combination of a speech masker in the listener's L2 and L1 ought to produce different decreases in the identification performance of L2 phonemes. A comparison of the decreases should reflect the degree of the contribution of the L1-based phonetic processing. The L1 masker should lead to a greater decrease in identification performance for a listener whose processing is high L1-based than for a listener whose processing is less L1-based due to interference with L1-based phonetic processing.

Three steps were taken in this study with the following aims. 1) To examine whether or not the speech-in-speech masking paradigm is effective to study interference with the processes underlying phoneme perception. 2) To examine whether or not L1 maskers decrease identification performance as we assumed. 3) To examine whether or not the decrease in the identification performance for L2 phonemes differs depending on the age group. Although L2 was used as a target in steps 2 and 3, L1 was used as a target in step 1. The reason for using L1 in step 1 was that a masker needs to have higher similarity with the target as well as greater consistency with the listener's L1 than with the other masker so that the masker interferes with both processes underlying phoneme perception (the perception of acoustic features and phonetic processing) and so that it produces larger decreases than the other masker.

Experiment 1, in step 1, examined whether or not the interference increased as the similarity of underlying processes for the target and masker increased. Experiment 2, in steps 2 and 3, examined whether or not the degree of interference differed depending on putative degrees of L1-based processing and on age.

## 2. Experiment 1

Experiment 1 was designed to examine whether or not L2 and L1 speech maskers produce different decreases in identification performance due to cases of interference with the underlying processes of phoneme perception. A combination of the listener’s L1 masker and L1 target phoneme was used to cause simultaneous interference with processes of perceiving acoustic features and phonetic encoding. A combination of the listener’s L2 masker and L1 target phoneme was used for a comparison.

### 2.1. Methods

#### 2.1.1. Stimuli

The target for Japanese listeners included four-mora Japanese words. Seventy words spoken by a male speaker (mya) were selected from a word-familiarity database (FW03) [12] by considering phonetic balances (e.g., “na.ki.yo.ri”; see Appendix). The words were selected from the highest and lowest familiarity rank to include a high range of familiarity. The target for American English listeners included English words contrasting /r/-/l/. Forty-eight minimal pairs (96 words) were selected from the stimuli used in the perceptual training [6]. They consisted of four phonetic environments, FP, FC, IP, and IC (e.g., FP: “bare–bail,” FC: “cold–cord,” IP: “rate–late,” and IC: “crock–clock”; see Appendix).

Each masker consisted of two-talker competing speech of sentences. The English utterances were randomly drawn from a TIMIT speech database [13], and the Japanese utterances were randomly drawn from “ATR 503 Sentence” B-set [14]. All utterances were derived from male talkers. Each masker was generated by summing two utterances (a different pair of talkers and sentences) with normalized peak intensity. Three maskers were generated per target, and one of them was presented to a participant.

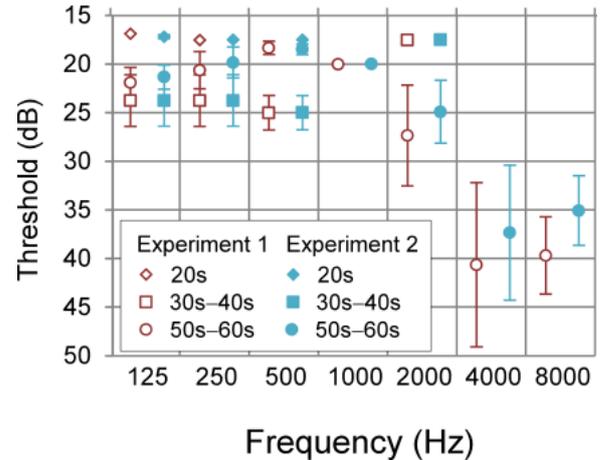
The targets were normalized for the peak intensity, and then the target (signal) and the masker (noise) were combined such that the signal-to-noise ratio (SNR) in the overlapped region was  $-6$  dB to produce the same amount of energetic masking. The masker started 800 ms earlier and ended 800 ms later than the target. Targets without maskers were prepared for a non-masker condition. The targets were identical for all masker conditions.

#### 2.1.2. Participants

A total of twenty-three native speakers of Japanese and eight native speakers of American English partici-

**Table 1** Grouping of participants in Experiment 1 into first language and age groups.

Group	Age (mean)	N
Japanese		
20s	22–27 (24.4)	7
30s–40s	35–41 (38.8)	8
50s–60s	55–68 (64.0)	8
American English		
	28–68 (42.3)	8



**Fig. 1** Means and standard errors of thresholds of participants in Experiment 1 (Japanese listeners) and Experiment 2 who did not pass a pure-tone hearing screening at 15 dB hearing level.

ated in Experiment 1 (Table 1). The Japanese participants were grouped into three age groups as identified in Table 1. No participant reported a history of hearing or speech disorders. A pure-tone hearing screening at 15 dB hearing level was administered at 250–8000 Hz only to Japanese listeners. The threshold at the frequency was measured for the participants who did not pass the screening at any of the frequencies. Fig. 1 shows the participants’ mean thresholds. The thresholds of 50s–60s were poorer than those of 20s or of 30s–40s for high frequencies. No Japanese listener had the experience of living abroad for more than three months. Four American English listeners had been living in Japan for 3–13 years.

#### 2.1.3. Procedure

Participants were tested in a sound-proof booth with a background noise of less than A-weighted 21 dB sound pressure level. All tests were administered on a personal computer (ThinkPad x121e) to present stimuli and to collect responses. Stimuli were presented from the personal computer through an audio interface (Fostex HP–A3) and headphones (AKG K272HD) diotically at ap-

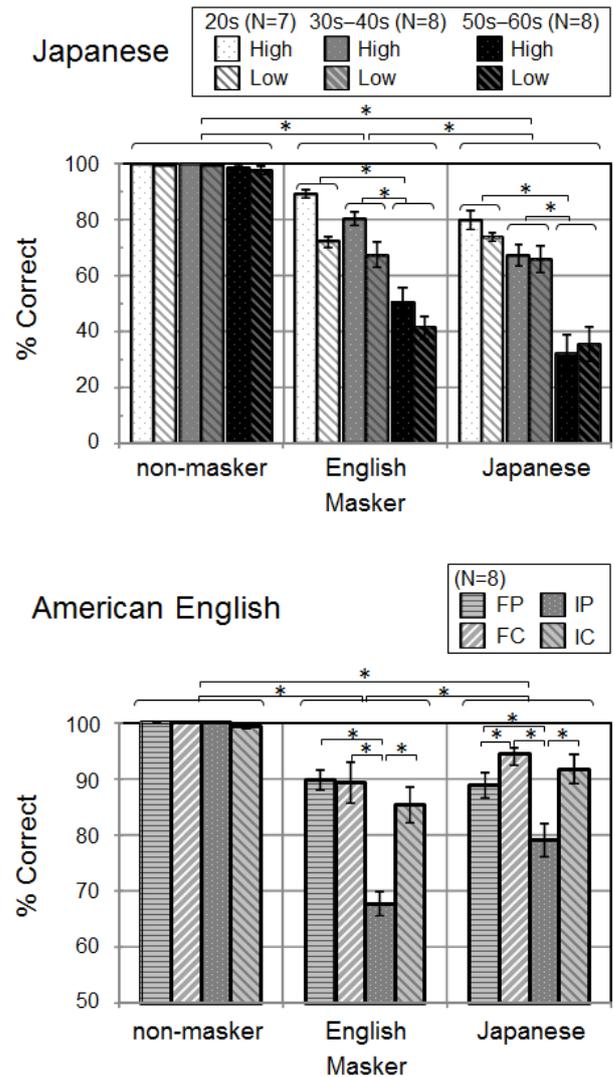
proximately A-weighted 67 dB sound pressure level. Participants were allowed to adjust the sound volume to their own comfortable level if needed. The participants were permitted to replay each stimulus up to ten times.

For Japanese targets, the participants heard a stimulus and were required to type what the participants heard in katakana characters (Japanese phonetic symbols). The task for English targets was a two-alternative-forced-choice (2AFC) identification task. On each trial, the participants heard a stimulus and were required to identify the stimulus by clicking one of the minimal pair displayed on a PC screen.

The stimuli were presented in random order, blocked by the type of masker (non, English, and Japanese). The order of the without-masker (non) and with-masker (English and Japanese) differed between Japanese and English targets. Because Japanese targets were tested with open-ended questions, a high possibility existed that the listener remembered the words and relied on memory during the subsequent blocks if the without-masker (non) condition was presented first. To reduce this influence, the with-masker (English and Japanese) conditions were the first and second block, and the without-masker (non) condition was the third block for Japanese targets. However, English targets were tested with a closed-ended question (2AFC). The order effects of without- and with-maskers can be considered to be small with respect to memorization. The order of the masker conditions was the same as that with Japanese listeners in Experiment 2. The without-masker (non) condition was the first block, and the with-masker (English and Japanese) conditions were the second and third blocks. The order of the presentation of English and Japanese maskers was counterbalanced to reduce order effects for both Japanese and American English targets.

## 2.2. Results

Japanese listeners' answers were analyzed for mora intelligibility scores by calculating the percentage of identified mora. Means and standard errors are shown in the top panel in Fig. 2. A repeated measures ANOVA was conducted with the arcsine-transformed scores as the dependent variable, with masker (non vs. English vs. Japanese) and familiarity (high vs. low) as the within-subject factors, and with age group (20s vs. 30s–40s vs. 50s–60s) as the between-subjects factor. The analysis revealed a significant main effect of masker ( $F(2, 40) = 455.32, p < .01$ ). A pairwise comparison with Bonferroni correction showed differences across all masker levels ( $p < .05$ ) such that the performance, as measured



**Fig. 2** Means and standard errors of identification performance for Japanese targets by Japanese listeners (top panel) and for English targets by English listeners (bottom panel).

by the percentage-correct scores, with the Japanese masker was the lowest. The analysis also revealed other main effects of familiarity ( $F(1, 20) = 29.54, p < .01$ ) and age group ( $F(2, 20) = 23.35, p < .01$ ) as well as significant interactions between masker and age group ( $F(4, 40) = 16.03, p < .01$ ) and between masker and familiarity ( $F(2, 40) = 24.42, p < .01$ ). The pairwise comparison with Bonferroni correction of the main effects showed that the scores for high familiarity were higher than ones for low familiarity ( $p < .05$ ), and that the scores of 50s–60s were lower than those of other age groups ( $p < .05$ ). The pairwise comparison with Bonferroni correction of the interactions showed that the scores of 50s–60s were lower than the ones of 20s and 30s–40s with both English and Japanese maskers and that the scores for high familiarity were higher than the

ones for low familiarity in non and English masker conditions ( $p < .05$ ), while significant differences did not exist between familiarities with the Japanese masker.

The identification performance of American English listeners on the American English targets are shown in the bottom panel in Fig. 2. A repeated measures ANOVA was conducted with the arcsine-transformed scores as the dependent variable and with masker (non vs. English vs. Japanese) and phonetic environment (FP vs. FC vs. IP vs. IC) as the within-subject factors. The analysis revealed a significant main effect of masker ( $F(2, 14) = 102.54, p < .01$ ). A pairwise comparison with Bonferroni correction showed differences across all masker levels ( $p < .05$ ) such that the scores in the English masker was the lowest. The analysis also revealed the main effect of phonetic environment ( $F(3, 21) = 16.47, p < .01$ ) as well as a significant interaction between masker and phonetic environments ( $F(6, 42) = 6.22, p < .01$ ). The pairwise comparison with Bonferroni correction of the main effect showed that the scores for IP were lower than the ones for other phonetic environments ( $p < .05$ ). The pairwise comparison with Bonferroni correction of the interactions showed that the scores for IP were lower than the ones of other phonetic environments in both of English and Japanese maskers ( $p < .05$ ).

Overall, both Japanese and American English listeners perceived L1 phonemes as being less intelligible in the L1 maskers compared to ones in the L2 maskers.

### 2.3. Discussion

Experiment 1 showed the L1 maskers had larger effects than the L2 maskers on the decrease in identification performance for the L1 phonemes. These results are in line with the previous findings dealing with speech-in-speech perception [9,10]. The L1 maskers possess high target-masker similarity and consistency with the listener’s L1. In contrast, the L2 maskers possess low target-masker similarity and inconsistency with the listener’s L1. These similarities and consistencies may interfere with processes underlying phoneme perception. Thus, the fact that L1 maskers produced a large decrease in identification performance supports the regard that the interference (i.e., the decrease in identification performance) increases in speech-in-speech perception as the similarity of underlying processes for the target and masker increases.

The differences between familiarities suggest that lexical information is also involved in the results. It might be questionable whether or not the Japanese listeners processed the targets at the phonetic level because the targets words were not presented as a phonetic con-

trast like English targets. However, because half of the Japanese targets were drawn from the lowest familiarity words, which contain very poor lexical information, the listeners presumably made their judgments based on a sequence of phonemes rather than relying on lexical information. The differences between age groups suggest that the maskers interfered with phoneme perception to a greater degree for the older age group than for younger ones. This agrees with previous findings examining speech perception in noises [15]. Even though differences were evident among age groups, the results showing that the Japanese maskers decreased the identification performance of the Japanese listeners regardless of the age group imply that the maskers’ interference with the processes was not dependent on the age group.

In summary, the findings demonstrated that the masker, which was assumed to interfere with the processes underlying phoneme perception, produced a large decrease in identification performance, indicating the effectiveness of the speech-in-speech paradigm.

## 3. Experiment 2

Experiment 2 was designed to examine whether or not L1 maskers decrease identification performance depending on the putative degree of L1-based phonetic processing and whether or not differences can be found in the effects of the masker among age groups. First, the differences in the effects of the maskers between putative L1-based processing of L2 phonemes were examined. Then, the differences among age groups and phonetic environments of /r-/l/ were examined.

### 3.1. Methods

#### 3.1.1. Stimuli

Two sets of English minimal pairs were prepared. The sets differed in “goodness of fit” to the associated Japanese categories. One was “Good-fit,” and the other was “Poor-fit.” The “Good-fit” set consisted of phonetic contrasts that resemble Japanese ones (e.g., /f/-/s/). Eight minimal pairs (sixteen words) were prepared (e.g., “defend–descend”; see Appendix), and each word was pronounced twice, so the set consisted of thirty-two utterances. The “Poor-fit” set consisted of ninety-six words contrasting /r-/l/ that were used in Experiment 1, which do not exist as a contrast in the Japanese phonological system (e.g., “rate–late”; see Appendix). Utterances by a male speaker were used in each set. Maskers were generated and combined with the targets in the same way as those in Experiment 1. Targets without maskers were prepared for a non-masker condition. The targets were identical for all masker conditions.

**Table 2** Grouping of participants in Experiment 2 into age groups

Group	Age (mean)	N
20s	22–29 (24.8)	12
30s–40s	35–44 (39.3)	9
50s–60s	51–68 (60.1)	13

### 3.1.2. Participants

A total of thirty-four native speakers of Japanese participated in Experiment 2. The participants were grouped into three age groups as defined in Table 2. One participant in the 50s–60s group was not able to have a test of “Good-fit.” The participant only had a test of “Poor-fit.” No participant reported a history of hearing or speech disorders. The same hearing screening and measurement of the thresholds were performed as those in Experiment 1. The mean thresholds of 50s–60s were poorer than those of 20s or of 30s–40s for high frequencies (see Fig. 1). Language background questionnaires were used to obtain their language experience. Table 3 shows the basic data of English learning. All participants had formal classroom learning from junior high school to senior high school. No one had the experience of living abroad for more than three months.

### 3.1.3. Procedure

Participants were tested in the same environment as the one in Experiment 1. At the beginning of the experiment, each participant was given stimulus familiarization tasks. One pair of each phonetic environment of /r-/l/ was selected for the familiarization tasks. The participant heard the stimuli repeatedly and then had a 2AFC task with correct/incorrect feedback. Following the familiarization tasks, the participant was tested the same as the English listeners in Experiment 1 were. The stimuli were presented in random order, blocked by the type of target (“Good-fit” and “Poor-fit”) first and by the type of masker (non, English, and Japanese) second. In order that the listeners could familiarize themselves with the contrast, learn criteria to identify the contrast, and use the criteria throughout all masker conditions constantly, the masker condition was presented in the

following order, with the non-masker condition being the first block and the English and Japanese masker conditions being the second and third blocks. The presentation order of the English and Japanese masker conditions was counterbalanced.

## 3.2. Results

Identification performance for “Good-fit” was calculated for each masker condition and participant using percentage-correct scores. To investigate the effects of the phonetic environment, scores for “Poor-fit” were calculated for each masker condition, participant, and phonetic environment. When a participant had a score at or below chance (Good-fit: 62.5 %, Poor-fit: FP 60.7 %, FC 65 %, IP 62.5 %, IC 62.5 %) in the non-masker condition, the participant’s scores in all masker conditions were eliminated from the following analyses. This was because being below chance in the non-masker condition indicated that the participant could not identify the contrast well, so the participant’s decreased or increased identification performance in maskers did not necessarily indicate masking effects occurred. It should be noted that, in case of “Poor-fit,” the participant’s scores were eliminated only for the corresponding phonetic environment. The numbers of participants used in the following analyses for “Poor-fit” were as follows: 20s: 10 (FP), 9 (FC), 6 (IP), and 4 (IC); 30s–40s: 7 (FP), 3 (FC), 4 (IP), and 2 (IC); and 50s–60s: 9 (FP), 6 (FC), 4 (IP), and 1 (IC). No eliminations were made for “Good-fit.”

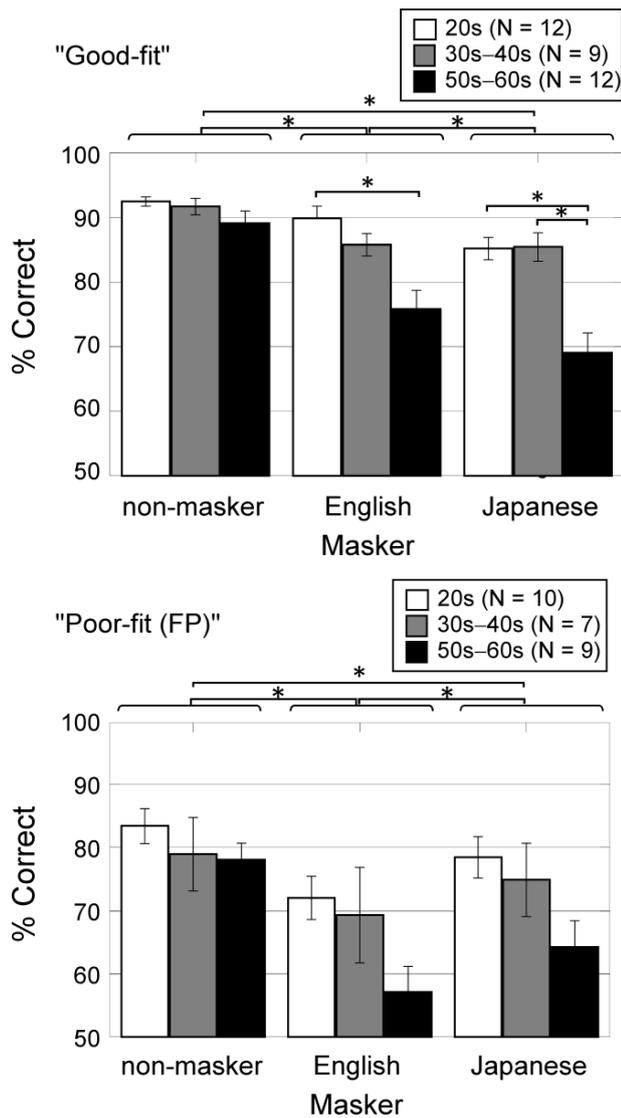
### 3.2.1. Differences between types of “goodness of fit”

The scores for “Good-fit” and for final singleton (FP) of “Poor-fit” are shown in Fig. 3. The results of “FP” are shown as a representative of “Poor-fit” to compare the results between types of “goodness of fit.” Other phonetic environments of “Poor-fit” are not shown here due to a lack of sufficient samples for the comparison caused by the elimination rule.

A repeated measures ANOVA was conducted with the arcsine-transformed scores as the dependent variable and with masker (non vs. English vs. Japanese) as the within-subjects factor and with age group (20s

**Table 3** English learning experience of participants in Experiment 2

	20s	30s–40s	50s–60s
Age of initial English learning (mean)	11.6	11.0	12.6
Years of formal English classroom learning (mean)	9.3	6.4	7.7
Number of participants who had English learning other than formal classroom	5	1	4



**Fig. 3** Means and standard errors of identification performance for “Good-fit” English phonetic contrast (good fit to the associated Japanese phonetic categories) by Japanese listeners (top panel) and for the final singleton (FP) of the “Poor-fit” English phonetic contrast (poor fit to the associated Japanese phonetic categories) by Japanese listeners (bottom panel).

vs. 30s–40s vs. 50s–60s) as the between-subjects factor, separately for “Good-fit” and “Poor-fit.” As for “Good-fit,” the analysis revealed a significant main effect of masker ( $F(2, 60) = 39.61, p < .01$ ). A pairwise comparison with Bonferroni correction showed differences across all masker levels ( $p < .05$ ) such that the scores in the Japanese masker were the lowest. The analysis also revealed the main effect of age group ( $F(2, 30) = 9.18, p < .01$ ) as well as a significant interaction between the masker and age group ( $F(4, 60) = 6.06, p < .01$ ). The pairwise comparison with Bonferroni correction of the main effect showed that the scores of 50s–60s were lower than the ones of other age groups ( $p < .05$ ).

The pairwise comparison with Bonferroni correction of the interactions showed that the scores of 50s–60s were lower than the ones of 20s for English maskers and were lower than the ones of 20s and 30s–40s for Japanese maskers ( $p < .05$ ). For the “Poor-fit (FP),” the analysis revealed a significant main effect of masker ( $F(2, 46) = 29.93, p < .01$ ). The pairwise comparison with Bonferroni correction showed differences across all masker levels ( $p < .05$ ) such that the scores for English maskers were the lowest. The main effect of age group and the interaction between these two factors were not significant.

A large decrease in identification performance for “Good-fit” English phonetic contrasts was produced by Japanese maskers that possess a high consistency with the listener’s L1. In contrast, a large decrease in the identification performance for the “Poor-fit” English phonetic contrast (FP) was produced by English maskers that possess high target–masker similarity. The results for the “Poor-fit (FP)” were in line with the results obtained in previous research [9]. However, the results for “Good-fit” were not in agreement with the research but rather similar to the results of Experiment 1. In summary, the relative effects of the maskers were dependent on the “goodness of fit” of the target phonetic contrasts.

### 3.2.2. Differences among age groups and phonetic environments of /r/-/l/

To examine whether or not the maskers had different effects among age groups and phonetic environments, the results for the “Poor-fit” contrast were further analyzed using phonetic environment including the environments excluded in Section 3.2.1. The relative effects of the Japanese and English maskers were evaluated in terms of the “intelligibility difference” of the arcsine-transformed scores in the Japanese and English maskers. Positive and negative values indicate that the scores were lower and higher for the Japanese than for the English maskers, respectively, implying smaller and larger effects of the English masker. For each participant and for each phonetic environment of the “Poor-fit” contrast, the intelligibility difference was calculated by subtracting the arcsine-transformed identification performance in the Japanese masker from that in the English masker. To consider the influence of perceptual pattern and difficulties identifying the phonetic contrast by native speakers of Japanese [8], the differences were pooled for each phonetic environment group, FP, FC and IP, IC. Histograms were generated to show the distributions of the intelligibility differences.

Table 4 displays the means and standard deviations

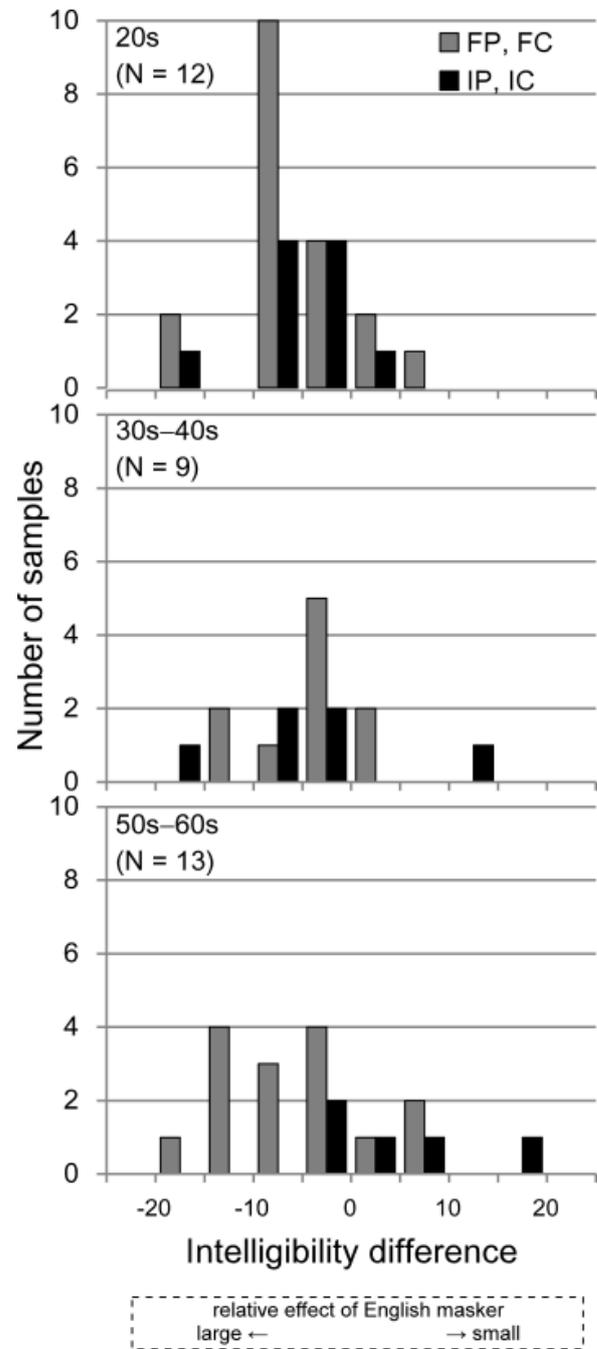
**Table 4** Means and standard deviations of intelligibility difference between English and Japanese maskers (arcsine transformed score) for each L1, age group, and phonetic environment group (FP, FC and IP, IC)

Group	FP, FC	IP, IC
Japanese		
20s	-5.28 (6.87)	-4.48 (6.35)
30s-40s	-3.16 (7.44)	-3.68 (8.60)
50s-60s	-4.09 (7.53)	5.32 (8.67)
American English		
	-1.69 (9.94)	-7.13 (6.20)

of the intelligibility differences for each L1, age group, and phonetic environment group (FP, FC and IP, IC). The values of American English were calculated from the results of Experiment 1. The means for FP, FC were under 0 for all L1 and age groups; however, the means for IP, IC increased as the age of the Japanese listeners increased.

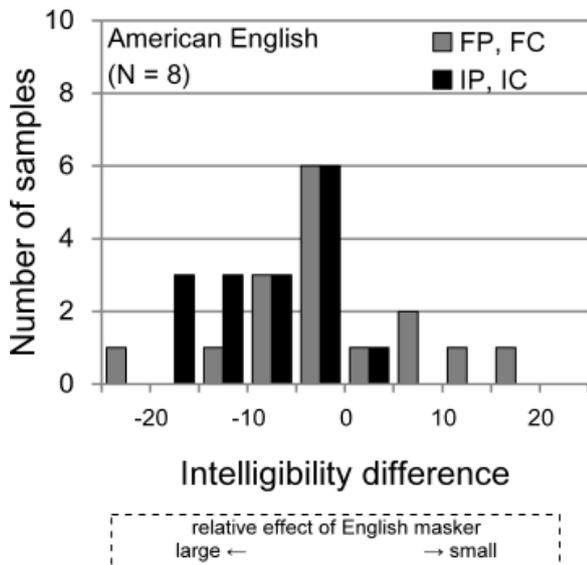
Fig. 4 displays the distributions of the intelligibility differences for each age group of Japanese listeners. An ANOVA was conducted with intelligibility difference as the dependent variable and with age group (20s vs. 30s-40s vs. 50s-60s) as the between-subjects factor for each phonetic environment group (FP, FC vs. IP, IC). The analysis revealed a significant main effect of age group ( $F(2, 59) = 3.18, p < .05$ ). A pairwise comparison with Bonferroni correction showed differences across all age group levels such that the intelligibility differences of 50s-60s were larger than the ones of 20s. A marginal interaction was shown between age group and phonetic environment group ( $F(2, 59) = 2.54, p < .09$ ). The number of samples of participants in 50s-60s was relatively small in the analysis, yet Fig. 4 suggests the group had different distributions between FP, FC and IP, IC. Though only a marginal interaction was found, the mean values and the distributions suggest that the differences between age groups might be in IP, IC, rather than in all phonetic environments. To summarize, the effects of the English maskers were large on the perception of /r/-/l/ in FP, FC for all age groups and were small on the perception of /r/-/l/ in IP, IC for participants in their 50s-60s.

Fig. 5 displays the distribution of American English listeners tested in Experiment 1. A repeated ANOVA was conducted with intelligibility difference as the dependent variable and with phonetic environment group (FP, FC vs. IP, IC) as the within-subjects factor. The analysis revealed a significant main effect of phonetic environment group ( $F(1, 15) = 5.00, p < .05$ ). The pairwise comparison with Bonferroni correction showed



**Fig. 4** Japanese listeners' intelligibility difference between English and Japanese maskers for each age group. The x-axis for each histogram represents the difference calculated by subtracting the arcsine-transformed score in the Japanese maskers from that in the English maskers (English minus Japanese). The y-axis represents frequency per age group. The frequency is broken down into FP, FC and IP, IC.

differences across all phonetic environment group levels such that the intelligibility differences for FP, FC were larger than ones for IP, IC. The means of American English listeners were under 0 both for FP, FC and IP, IC, but the distributions were different. These can be summarized as larger effects of the English masker regard-



**Fig. 5** American English listeners' intelligibility difference between English and Japanese maskers. The x-axis represents the difference calculated by subtracting the arcsine-transformed score in the Japanese maskers from that in the English maskers (English minus Japanese). The y-axis represents frequency. The frequency is broken down into FP, FC and IP, IC.

less of phonetic environments, and there were differences in the size of the effects among phonetic environments.

Despite the target-masker similarity and consistency with the listener's L1 being identical over all phonetic environments and age groups of Japanese listeners, the relative effectiveness of the maskers was different. The relative effectiveness was equivalent among age groups for FP and FC; however, the effects of the English masker were relatively smaller in older adults for IP and IC.

### 3.3. Discussion

#### 3.3.1. Decreases in identification performance and degree of the contribution of the L1-based processing

As shown in Section 3.2.1., the decreases in the relative effects of the maskers were dependent on the "goodness of fit" of the target L2 phonetic contrasts to the listener's L1 phonetic categories. According to the principle where the listener's L1 interacts with L2 perception, putative high and less L1-based processing would emerge to identify "Good-fit" and "Poor-fit" L2 phonetic contrasts, respectively, because L1-based perception facilitates the perception of L2 phonetic contrasts when it works; in contrast, it inhibits that perception when it does not work well. The results could be summarized as follows. The Japanese masker produced a relatively large decrease in the identification performance

when high L1-based processing emerged. In contrast, the English masker produced a relatively large decrease in the identification performance when less L1-based processing emerged.

Not only in the perception of L1 phonemes, as shown in Experiment 1 (Fig. 2), a larger decrease in the identification performance of L1 maskers appeared even in the perception of L2 phonemes. This implies that not only the process of perceiving the acoustic features was interfered with by the acoustic similarity but also other processes (i.e. phonetic processing) were interfered with. The differences between "goodness of fit" imply the decrease could be associated with the interference with the L1-based phonetic processing. It showed that the degree of the L1-based phonetic processing can be estimated from the decreases in the identification performance of the maskers.

#### 3.3.2. Differences among age groups and phonetic environments, and L1-based perception for L2 phonemes

The results of the Japanese listeners showed that the effects of the Japanese maskers were equivalent among age groups for the final singleton and final cluster positions, but the effects increased in older adults for the initial singleton and initial cluster positions (Table 4, Fig. 4). Based on the interpretation of the decrease in the identification performance, the degree of the L1-based phonetic processing can be estimated from the decreases in the identification performance of the maskers (Section 3.3.1.); the findings imply that the adults in the age group of 50s–60s were more L1-based than the adults in other age groups in identifying /r-/l/. The differences among age groups likely are when the phonetic contrast is in the initial singleton and initial cluster positions.

The results were not in line with the initial expectations based on the perceptual pattern and difficulties to identify the sound by native speakers of Japanese [8] and training effect [6]. The results of these studies led to the expectation of intelligibility differences in the final singleton and final cluster positions being larger than the ones of the initial singleton and initial cluster, especially among older adults. Contrary to this expectation, the results demonstrated that the intelligibility differences in the final singleton and final cluster were smaller than the ones in the initial singleton and initial cluster. It is interesting to note that the intelligibility differences for American English listeners were different among phonetic environments. The listeners might use one of multiple redundant cues for speech perception in maskers; the differences among their L1 and phonetic environ-

ments might reflect the differences in their phonetic processing. The differences in the phonetic environment of the masking effects remain for further research.

In addition, the aging effects on the effects of maskers should be treated with caution. Older adults tend to have larger masker effects than younger adults do [15]. The oldest age group in the study indeed had identification performance that was lower than the ones of other age groups. However, the fact that the differences in the decreases in the identification performance among age groups for the final singleton (FP) of “Poor-fit (/r-/l/)” were not statistically different (Fig. 3, bottom panel) indicates that the aging effects were not severe enough to impair listening in the maskers in the study. In addition, the fact that the oldest age group decreased differently between the “goodness of fit” in the same pattern as other age groups (Fig. 3) indicates that the estimate of the degree of the L1-based processing can be supported.

In conclusion, the findings showed that the decreases in identification performance of speech maskers depended on the age group and the phonetic environment of /r-/l/. This implies that the perception of /r-/l/ was altered to be high L1-based in older adults, especially in the perception of the phonetic contrast in the initial singleton and initial cluster positions.

#### 4. General Discussion

The present study suggests that the perception of the phonetic contrasts by the old age group is higher L1-based than the one of the young age group, especially for the perception of the phonetic contrast in the initial singleton and initial cluster positions.

In addition to these findings, let us consider the possible explanation for the age-related decrease in the training effectiveness [6]. The explanation was that phonetic processing for the L2 phonemes was altered to be high L1-based phonetic encoding in adulthood, and it brought about an age-related decrease in the training effectiveness. Taken together with the results of this study and the training study, where high L1-based phonetic processing was suggested in this study, the participants in the perceptual training did not improve (i.e., older adults for the initial singleton and initial cluster positions). Similarly, where less L1-based phonetic processing was suggested in the present study, the participants in the perceptual training improved (i.e., young adults for all positions and older adults for the final singleton). These findings are supportive of the explanation because they indicate that perceptual differences may be involved in the differences in the perceptual training. The relationship between the perceptual differences and

training effectiveness needs to be investigated in future research.

#### 5. Conclusion

This study examined the dependence on age of interference with phoneme perception by L1 and L2 speech maskers. It was implied that the identification performance in the maskers decreased as the underlying processes for the target and masker increased. The findings support using the decreases in the identification performance to estimate the degree of the contribution of the L1-based phonetic processing. The perception of American English /r-/l/ may have been altered to be high L1-based in older adults of Japanese, especially when the phonetic contrast was in the initial singleton and initial cluster positions.

#### Appendix A.

[low familiarity] hya.ku.shyu.tsu ga.ra.yu.ki ka.wa.da.chi  
bi.ru.shya.na mi.zu.gu.shi ha.ko.se.ko to.me.ya.mak  
sa.su.pu.ro bo.te.fu.ri ku.ni.ta.mi ji.ri.da.ka shi.wa.ba.ra  
ma.chi.ya.ne chyo.ki.bu.ne te.tsu.mu.ji se.se.na.gi  
mo.to.go.me mo.no.no.gu do.bu.zu.ke jya.ku.jya.ku  
ta.ka.hi.mo na.ki.yo.ri ta.wa.re.me he.ra.zu.ke tsu.ke.za.shi  
ge.ba.ru.to so.ma.go.ya fu.ji.gi.nu ge.ki.ze.tsu ka.wa.ho.ne  
be.ni.ga.ra ha.na.mi.zo de.ko.mo.no kyo.ku.hi.tsu  
shi.ki.ri.shyo

[high familiarity] ka.ke.so.ba ga.ku.bu.chi ki.ki.to.ri  
kya.ra.me.ru ge.tsu.ma.tsu ko.re.ho.do go.ta.go.ta  
shi.na.mo.no ji.gu.za.gu yo.bi.su.te ka.za.ka.mi ga.ku.wa.ri  
ki.zu.mo.no shi.na.gi.re shya.bu.shya.bu chya.ru.me.ra  
do.ro.nu.ma ne.ko.ba.ba hi.to.ji.chi ho.ya.ho.ya ga.ni.ma.ta  
kya.ku.se.ki ke.shi.go.mu sa.ke.gu.se ze.tsu.me.tsu  
so.pu.ra.no to.mo.da.chi na.zo.na.zo ha.ba.yo.se pa.ri.pa.ri  
bi.shyo.nu.re yu.de.da.ko ka.be.ga.mi ko.ku.fu.ku pi.su.to.ru

#### Appendix B.

[Poor-fit (/r-/l/)] [FP] bare–bail boar–bowl dare–dale dire–  
dial fair–fail fear–feel hare–hail mare–mail mire–mile pare–  
pail roar–roll sore–soul tear–tail wire–while [FC] bird–  
build card–called cord–cold gourd–gold heart–halt mart–  
malt mired–mild sheared–shield tired–tiled wired–wield [IP]  
rate–late red–led rear–leer rude–lewd rice–lice rim–limb  
rhyme–lime rip–lip writ–lit rock–lock wrong–long rot–lot  
[IC] grew–glue pray–play pry–ply braze–blaze bruise–blues  
cram–clam crowd–cloud kraut–clout crown–clown groat–  
gloat prop–plop bracken–blacken

#### Appendix C.

[Good-fit] deep–keep hope–soap boat–boot get–got  
swimming–swinging defend–descend him–hip mad–man

#### Acknowledgments

A part of this work was supported by the Strategic Information and Communications R & D Promotion Programme (SCOPE; 131205001) of the Ministry of Internal Affairs and Communications (MIC), Japan.

## REFERENCES

- [1] E. H. Lenneberg: *Biological Foundations of Language*, (John Wiley & Sons, New York, 1967).
- [2] J. S. Logan, S. E. Lively and D. B. Pisoni, "Training Japanese listeners to identify English /r/ and /l/: A first report," *J. Acoust. Soc. Am.*, **89**, 874–886 (1991).
- [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," *J. Acoust. Soc. Am.*, **96**, 2076–2087 (1994).
- [4] A. R. Bradlow, D. B. Pisoni, R. Akahane-Yamada and Y. Tohkura, "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," *J. Acoust. Soc. Am.*, **101**, 2299–2310 (1997).
- [5] R. Kubo and R. Akahane-Yamada, "Influence of aging on perceptual learning of English phonetic contrasts by native speakers of Japanese," *Acoust. Sci. & Tech.*, **27**, 59–61 (2006).
- [6] R. Kubo, R. Akahane-Yamada and M. Akagi, "Effects of perceptual training on the ability of elderly adults of Japanese speakers to identify American English /r/-/l/ phonetic contrasts," Poster presented at the Acoustics 2012 Hong Kong, Proc. Acoustics 2012 Hong Kong, p. 124 (Abstract).
- [7] S. E. Lively, J. S. Logan and D. B. Pisoni, "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," *J. Acoust. Soc. Am.*, **94**, 1242–1255 (1993).
- [8] R. Komaki and Y. Choi, "Effects of native language on the perception of American English /r/ and /l/: A comparison between Korean and Japanese," Proc. 14th Int. Congr. of Phonetic Science, pp. 1429–1433 (1999).
- [9] S. Brouwer, K. J. Van Engen, L. Calandruccio and A. R. Bradlow, "Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content," *J. Acoust. Soc. Am.*, **131**, 1449–1464 (2012).
- [10] L. Calandruccio, S. Brouwer, K. J. Van Engen, S. Dhar and A. R. Bradlow, "Masking release due to linguistic and phonetic dissimilarity between the target and masker speech," *Am. J. Audiol.*, **22**, 157–164 (2013).
- [11] K. J. Van Engen and A. R. Bradlow, "Sentence recognition in native- and foreign-language multi-talker background noise," *J. Acoust. Soc. Am.*, **121**, 519–526 (2007).
- [12] S. Amano, K. Kondo, Y. Suzuki and S. Sakamoto, "Japanese speech dataset for familiarity-controlled spoken-word intelligibility test (FW03)," NII Speech Resources Consortium (2006).
- [13] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," <http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S1> (1993).
- [14] NII-SRC, "ATR 503 Sentences," <http://research.nii.ac.jp/src/en/ATR503.html>.
- [15] R. Rajan and K. E. Cainer, "Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational maskers," *Neuroscience*, **154**, 784–795 (2008).

## Cover Letter for Acoustical Science and Technology

- (1) **Title of paper**  
Dependence on age of interference with phoneme perception by first- and second-language speech maskers
- (2) **Full name(s) of author(s)**  
Rieko Kubo<sup>1,\*</sup>, Masato Akagi<sup>1,†</sup> and Reiko Akahane-Yamada<sup>2,‡</sup>
- (3) **Affiliation(s)**  
<sup>1</sup>School of Information Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, 923-1292 Japan  
<sup>2</sup> ATR Intelligent Robotics and Communication Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan
- (4) **Approximately five keywords**  
Speech perception, Second language, Phonetic processing, Aging effect, Influence of first language, Speech-in-speech masking, Informational masking
- (5) **PACS number**  
43.71.Hw, 43.71.Es, 43.66.Dc, 43.71.Lz,
- (6) **Short running title**  
R. KUBO *et al.*: DEPENDENCE ON AGE OF PHONEME PERCEPTION IN MASKERS
- (7) **Category of article:**  
PAPER
- (8) **Mailing address**  
石川県能美市旭台 1-1  
北陸先端科学技術大学院大学  
久保理恵子
- (9) **Classification**  
Psychological and Physiological Acoustics
- (10) **Number of pages**  
TEXT: 11    FIGURES: 5    TABLES: 4