English Language & Literature Teaching, Vol. 11, No. 2 Summer 2005

The Effect of Audio and Visual Cues on Korean and Japanese EFL Learners' Perception of English Liquids*

Hyunsong Chung
(Korea National University of Education)

Chung, Hyunsong. (2005). The effect of audio and visual cues on Korean and Japanese EFL learners' perception of English liquids. *English Language & Literature Teaching, 11*(2), 135-148.

This paper investigated the effect of audio and visual cues on Korean and Japanese EFL learners' perception of the lateral/retroflex contrast in English. In a perception experiment, the two English consonants /l/ and /r/ were embedded in initial and medial position in nonsense words in the context of the vowels /i, a, u/. Singletons and clusters were included in the speech material. Audio and video recordings were made using a total of 108 items. The items were presented to Korean and Japanese learners of English in three conditions: audio-alone (A), visual-alone (V) and audio-visual presentation (AV). The results showed that there was no evidence of AV benefit for the perception of the /l/-/r/ contrast for either Korean or Japanese learners of English. Korean listeners showed much better identification rates of the /l/-/r/ contrast than Japanese listeners when presented in audio or audio-visual conditions.

[perception/audio and visual cues/Korean and Japanese EFL learners]

I. INTRODUCTION

A multiplicity of acoustic cues to phonetic features are contained in the speech signal, and this redundancy of information helps listeners in their native language to overcome degradation of the signal by noise, other voices or reverberation. Visual gestures can also provide cues to the place or articulation of vowels and the place and

* This paper contains one part of the unpublished co-work with Valerie Hazan, which investigates the effect of language background and visual salience on the use of visual cues in the perception of non-native contrasts. The author revised his own work for this publication.

manner of articulation of consonants (e.g. Summerfield, 1983). The relative perceptual weighting of visual and auditory cues does appear to vary across languages. Visual cues are less informative in some languages, and this leads listeners to primarily direct their attention to auditory cues. Infants do display sensitivity to visual information, as they prefer video presentations of talkers with congruent rather than incongruent audio and visual channels. However, there is also evidence that the use of speechreading cues develops with age, as children aged 6 to 10 years are less influenced by visual information in their judgments of consonant identity than are adult learners (Massaro, Thompson, Barron, & Laren, 1986). This gradual development in the use of speechreading cues mirrors increasing sensitivity to certain acoustic cues over the first 10 years of life (Hazan and Barrett, 2000).

Current models of L2 acquisition only consider auditory cues to phonemic contrasts, and it must be established whether it is also possible for learners of a language to attune to the visual cues marking phonemic contrasts. Few studies have investigated L2 learners' use of visual information. One such study investigated the perception of a wide range of English consonants and vowels presented in background noise in auditory and audiovisual conditions by Spanish-L1 talkers and native controls (Ortega-Llebaria, Faulkner, & Hazan, 2001).

The aim of this study was to evaluate Korean and Japanese EFL learners' sensitivity to visual cues for English phonemic contrasts that do not occur in their native phoneme inventory. The perception of these contrasts was tested with learners of English from two different language backgrounds in order to evaluate the effect of the native phoneme inventory on the acquisition of auditory and visual cues to the contrasts.

This study investigated the perception of the lateral/retroflex contrast in English between the phonemes /l/ and /r/. This contrast has been extensively used in studies of L2 speech perception as it is difficult for Korean and Japanese learners of English to acquire due to the different phonological status of this contrast in these languages (Kim, 2004; Im & Ahn, 2002). In modern standard Seoul Korean, a liquid is phonetically realized as a [l], [4] and [n]. These allophones are in complementary distribution. Liquids do not appear in word-initial positions except in recent loan words where the liquid becomes a flap (Kang, 1999). In intervocalic positions, a liquid becomes a flap [4], while it becomes a lateral in geminates in intervocalic positions.

In Japanese, there is a single liquid within the Japanese phoneme inventory. Although phonetic variation in the production of this phoneme occurs across talkers, this phoneme (Japanese /r/) is most often described as an alveolar flap [4]. The Japanese phoneme /r/ occurs at syllable onset only, in word-initial or word-medial position. The English phonemes /l/ and /r/ tend to be assimilated to the native alveolar flap, leading to problems in the discrimination and identification of these English phonemes. A number of studies have shown higher identification rates for English /r/ than /l/ for Japanese-L1 learners (Bradlow, Pisoni, Yamada, & Tohkura, 1997). The /l/-/r/ contrast is particularly difficult for Japanese learners because though it primarily marked by differences in third formant transitions to which Japanese listeners are not particularly sensitive to, they tend to give greater weight to more variable secondary cues such as F2 transition (Yamada, 1995; Gordon, Keyes, & Yung, 2001).

There is therefore the expectation that the English /l/-/r/ contrast might be easier for Korean-L1 learners than Japanese-L1 learners in intervocalic position but that in initial position, Korean-L1 learners could have greater perceptual difficulties. These effects were indeed found in a study of the perception of the /l/-/r/ contrast in Australian English by Korean-L1 and Japanese-L1 learners (Ingram & Park, 1998). For the /l/-/r/ contrast, this study hypothesized poor use of visual gestures by both groups.

II. RESEARCH DESIGN

1. Speech Material

The two English consonants /l/ and /r/ were embedded in initial and medial position in nonsense words in the context of the vowels /i, a, u/. Because different predictions are made in terms of the perception of /l/-/r/ in clusters according to L1-background, both singletons and clusters were included: the consonants were presented as singleton or cluster with the initial consonant of the cluster being /k/ and /f/ and appeared in the structure CV, CCV, VCV and VCCV.

2. Talkers and Recording Procedure

A phonetically-trained female speaker of south-eastern British English recorded the test items. Currently, a mixture of non-regional and local south-eastern English pronunciation and intonation is considered a new standard British English. It is also called ‘Estuary English’. So in this experiment, the south-eastern British English speaker was used for the recording. Video recordings were made in a soundproof room, with the talker's face set against a blue background and illuminated with a key and a fill light. The talker's head was fully visible within the frame. Video recordings were made to a Canon XL-1 DV camcorder. Audio was recorded from a Bruel and Kjaer type 4165 microphone to both the camcorder and to a DAT recorder. The video channel was digitally transferred to a PC and time-aligned with the DAT audio recording, which was of higher quality than the audio track of the video. Video clips were edited so that the start and end frames of each token showed a neutral facial expression. Stimuli were down-sampled once the editing had been completed (250*300 pixels, 25 f/s, audio sampling rate 22.05 kHz). Three items were produced for each consonant in each syllabic and vowel context (27 initial /l/ and /r/, 27 medial /l/ and /r/), yielding a total of 108 items.

3. Listeners

This study involved 42 Korean learners of English and 78 Japanese learners of English. All of the Korean-L1 listeners were university students of D University and tested in Korea. These learners were at a lower- to lower-intermediate level of English proficiency. The median number of years of English study was 10 years (range: 6 to 18 years) but with little focus on spoken English. Six of the students had spent more than 6 months in an English-speaking country.

Of the Japanese-L1 listeners, 42 were university students of K University and tested in Japan, 20 students were attending a Summer Course in Phonetics at a university in London, 9 were recruited from a School of English in London and 7 were students of a pre-academic language course at the university. These learners were at a lower- to lower-intermediate level of English proficiency. The median number of years of English study was 8 years (range from 2 to 20 years) but with little focus on spoken English. Only three of the students for whom information was obtained spent more than 6 months in an English-speaking country. Students reported normal hearing and normal or corrected vision.

A control group of 12 native English listeners judged the test items in the video alone condition to check the difficulty of the items.

4. Experimental Task

A closed-set identification task was built using the CSLU toolkit software (Cole, 1999), and a conversational agent (Massaro, 1998) was used to explain the task to the listener and to give general feedback on the percentage of correct responses at the end of each section of the test. The conversational agent is a 3-D model of a talking face with accurate articulatory movements and facial expressions. There were three test conditions: (1) Video alone presentation (V); (2) Audio alone presentation (A) and (3) Audiovisual presentation (AV), with two blocks of 108 items presented per condition. Each listener therefore heard 108 repetitions of each consonant (across vowels and positions) in each test condition. The order of items was randomized within each block. Two orders were used for the presentation of the three conditions: AV, A, V or A, AV, V, and the two orders were counterbalanced across listeners. Items were presented to both ears at a comfortable listening level via headphones.

III. RESULTS

1. Group Results

The percentage of correct /l/ and /r/ responses was calculated. For Japanese-L1 learners, mean /l/-/r/ identification was 61.5% (s.d. 13.7) for the A condition, 63.4% (s.d. 14.8) for the AV condition and 55.8 % (s.d. 8.7) for the V condition. For the Korean-L1 learners, mean /l/-/r/ was 88.3% (s.d. 14.3) for the A condition, 87.9% (s.d. 14.6) for the AV condition and 63.1% (s.d. 8.7) for the V condition. Scores were then converted to the signal detectability measure d-prime (d'), which is calculated as the z-value of the correct response rate minus that of the incorrect response rate for one of the consonants. For the Japanese group, because some learners were tested in the UK and others in their home country, a repeated-measures ANOVA was run to see whether the place of testing affected performance. As this effect was not significant, the data was grouped per language background for the main analyses. The group of native controls was not tested on the A and AV conditions because of likely ceiling effects but achieved a score of 72.1% (d' of 1.31) on the V condition. Therefore, for native listeners, identification using lipreading alone is above chance.

A repeated-measures analysis of variance evaluated the within-subject effect of condition (A, AV, V) and the between-subject effect of L1 background. As suggested by the mean scores, the effect of L1 background was significant with Korean-L1 learners achieving higher scores than Japanese-L1 learners [F(1,118)=95.40; p<0.001]. The effect of test condition was strongly significant [F(2,236)=195.88; p<0.001]. Pairwise comparisons with Bonferroni adjustments showed that this was due to lower scores being obtained in the V condition than in the A and AV conditions, but there was no significant difference between the A and AV conditions. The condition * L1 background interaction was also significant [F(2,236)=93.48); p<0.001]. Examination of means suggests that this is due to scores for A and AV (but not V) conditions being higher for Korean than Japanese listeners.

To check that the difference in performance across L1 groups was not due to significant differences in amount of English instruction, statistics were rerun on data from a subgroup of listeners with a more narrow range of years of learning (7 or 8 years of exposure to English). This subgroup included 41 Japanese-L1 and 10 Korean-L1 speakers. The same effects were obtained for this subgroup as described above for the complete dataset.

The data was then examined to compare the two L1 groups in terms of the 'AV benefit' that they showed, i.e. the degree to which the combined channels provided more information than the baseline auditory condition. Rather than merely look at the difference in scores between A and AV conditions, which introduces a bias for subjects with A scores near ceiling, a measure of relative benefit can be taken which is the 'difference between A and AV scores relative to the amount of performance improvement possible given the subject's A scores (Sumby & Pollack, 1954). This relative benefit score is calculated as (AV-A)/(1-A) for A and AV scores expressed in terms of probability correct. This measure is useful for comparing across L1 groups given their difference in performance in the A condition. The mean AV benefit was 0.05 (s.d. 0.28) for the Japanese-L1 group and –0.16 (s.d. 0.73) for the Korean-L1 group. An AV benefit close to zero was therefore obtained for both groups and a univariate ANOVA failed to show any significant difference in AV benefit across language groups.

To check that ceiling effects were not masking the effect of AV benefit (in Korean-L1 learners, especially), statistics were run on the data for the lowest 50% of listeners when ranked on their AV performance for each L1-group. Mean d' for the two subgroups were as follows: Japanese-L1 group A 0.98, AV 0.65, V 0.68; Korean-L1 group A 2.02, AV 1.94 and V 0.45. Repeated-measures ANOVAs were run on this selected group. The effect of condition was significant [F(2,122)=79.12; p<0.001] and Bonferroni-adjusted comparisons tests showed that this was due only to performance in the V condition being significantly lower than for A and AV conditions. On average, listeners therefore do not show evidence of AV benefit as they are not identifying the /l/-/r/ contrast better in the AV than in the A condition. The condition*L1 group interaction was also significant [F(2,122)=76.04; p<0.001] which seems to be due to the higher identification rate for Korean-L1 listeners in the A and AV but not the V condition. The degree of relative AV benefit was again compared for the two groups. The means of -0.24 for the Korean-L1 group and -0.01 for the Japanese-L1 group were not significantly different. Therefore, even for this lower-performing group, there was no evidence of AV benefit for either group of listeners.

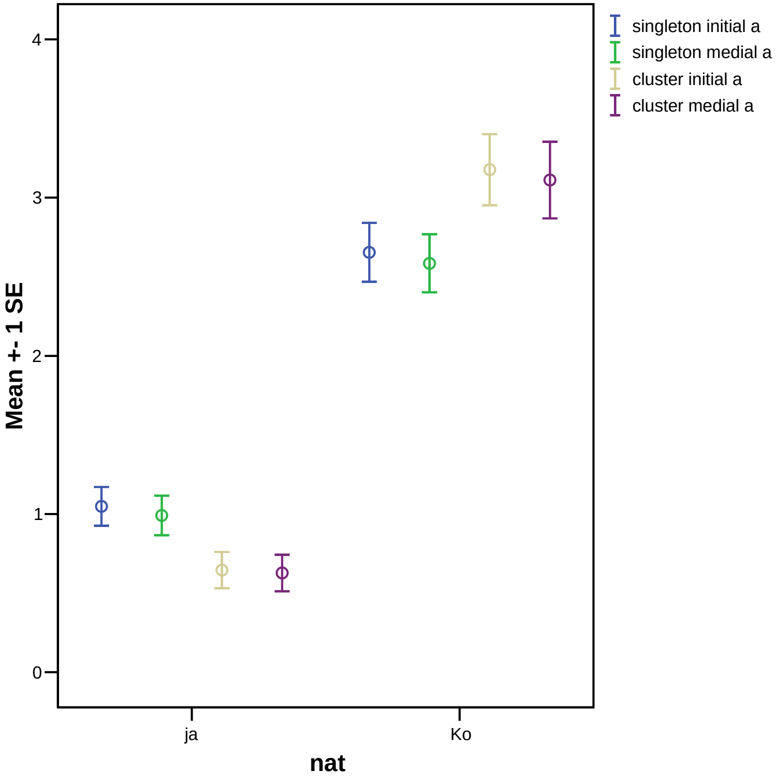
The effects of consonant, consonant position and syllable structure were examined in more detail (Figure 1), given the difference between the likely patterns of assimilation of L2 consonants to L1 categories in the two L1-groups. Identification scores were for each learner for the four following categories on consonants in each condition: singletons in word-initial position, singleton in word-medial position, clusters in initial position and cluster in word-medial position. According to Ingram and Park (1998), for Korean-L1 learners, the following order of difficulty was expected (from most difficult to easiest): word-initial singletons, word-initial clusters, word-medial singletons. Japanese-L1 learners are expected to find word-initial and word-medial singletons equally difficult to identify, and to have greatest difficulty with word-initial clusters. Analyses of variance for repeated measures were carried out on scores converted to d-prime.

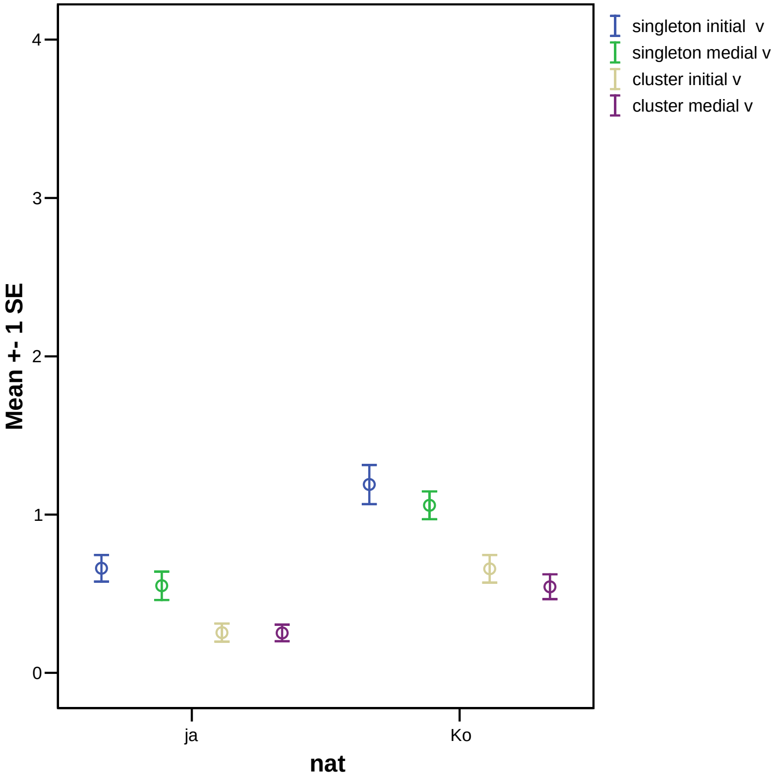
In the auditory condition, the main effect of syllable structure was not significant but there was a significant interaction between syllable structure and L1-group (F(1,118)= 47.35; p<0.001]. Observation of the data suggests that higher scores were obtained for clusters than singletons for Korean-L1 learners but not for Japanese-L1 learners. The only other significant effect was the three-way interaction between position, structure and L1-Group.

In the visual alone condition, the following effects were obtained. The main effect of syllable structure was significant with higher scores obtained for /l/-/r/ in singletons than in clusters [F(1,118)=105.77; p<0.001]. The main effect of position was also significant, with higher scores obtained for word-initial than for word-medial consonants [F(1,118)=5.79; p<0.02].

FIGURE 1

/l/-/r/ Identification (d’) for Word-initial and Word-medial Singletons and Clusters in the Auditory (A) and Visual (V) Conditions



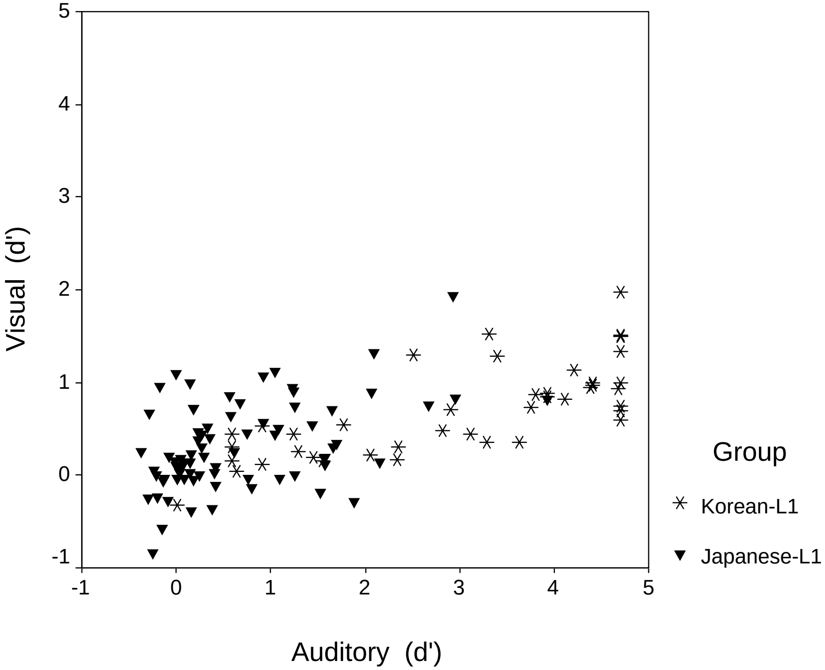


2. Individual Results

In Figure 2, a crossplot of /l/-/r/ perception in the A and V conditions is presented for individual Korean-L1 and Japanese-L1 listeners. The level of performance in Japanese-L1 listeners was typically lower than that of Korean-L1 listeners but there is evidence of some individuals with fairly good scores in the auditory condition but random performance on the basis of visual cues, and a few individuals with better use of visual than auditory cues.

The data was then examined to look at correlations between performance in the auditory and visual modalities. Scores in the A and AV conditions were strongly correlated, with a stronger correlation seen for the Korean-L1 group (r=0.967, N=42, p<0.0001) than for the Japanese-L1 group (r=0.884, N=78, p<0.0001). These strong correlations between conditions confirm the finding that adding visual information adds little advantage over the auditory condition alone. Pearson's correlations between performance in the A and V conditions were also significant for the Korean-L1(r=0.768, N=42, p<0.0001) and Japanese-L1 (r=0.477, N=78, p<0.0001) groups, suggesting that those learners which showed better consonant identification on the basis of acoustic information also showed better identification on the basis of visual information. Overall, although there is some evidence of above chance performance on the V condition in some learners at least, there is little evidence of AV integration.

FIGURE 2
/l/-/r/ Perception in the A and V Conditions
for Individual Korean-L1 and Japanese-L1 Listeners



IV. DISCUSSION

There was no evidence of AV benefit for the perception of the /l/-/r/ contrast for either Korean-L1 or Japanese-L1 learners of English, both for ‘selected’ and ‘unselected’ groups of listeners. Korean-L1 listeners show much better identification of the /l/-/r/ contrast than Japanese-L1 listeners when presented in audio or audio-visual conditions. Although Korean-L1 listeners also showed significantly higher /l/-/r/ identification in the visual condition than Japanese-L1 learners, there was a marked discrepancy in Korean-L1 learners between their use of auditory and visual information, whereas Japanese-L1 learners were generally poor in identifying the contrast using either modality. The poor use of visual information by learners, whatever their language background and level of proficiency is probably due to the low degree of visual salience of the contrast. It must be noted though that ‘high-performers’ do show significantly better use of visual cues (higher V score) than ‘low performers’, so there is some evidence of ‘attuning’ to visual cues with experience.

The effect of L1 background on the perception of the /l/-/r/ contrast is quite marked, with much higher levels of performance achieved by Korean-L1 learners. The effect of their greater degree of language instruction (median of 10 years of instruction for Korean-L1 learners versus 8 years for Japanese-L1 learners) cannot be discounted, but it must also be noted that this difference in general levels of performance between Korean and Japanese learners of English does mirror previous findings of Ingram and Park (1998), Park (1999) and Sakow and McNutt (1993). Ingram and Park (1998) suggested that Korean-L1 speakers do have some exposure to alveolar laterals, at least in word-medial position as these are produced in geminates (liquid in coda position followed by another in onset position), and that this exposure to a [4] – [ll] contrast in word-medial position might help in the acquisition of the English /r/-/l/ contrast, at least in word-medial position. In our data, in the auditory condition, this expected difference between performance in word-initial and word-medial position was not obtained. According to one of recent works by Lee & Lee (2004), in the Korean learners’ perception of English liquids (audio condition), word-final position was the most difficult and the word-initial cluster position was the easiest. Despite the absence of C+liquid clusters in either Korean or Japanese, Koreans showed significantly higher identification rates for consonants in either word-initial or word-medial clusters than in singletons, whereas Japanese-L1 learners showed poor performance in both positions and syllable structures. Korean syllable structure constraints do not allow syllable-initial consonant clusters. So when they pronounce English consonant clusters, they tend to insert central vowel [1] to the consonant cluster. This process also applies to the perception. So the Korean syllable structure constraints make English /l/ and /r/ in consonant clusters placed in intervocalic positions. This situation is the same as that in singletons in word-medial positions. Due to this reason, Korean learners seem to show higher identification rate for liquids in consonant clusters than Japanese learners. If liquids occur in a cluster such as Cl/Cr, then Korean learners can compare the sound quality between C, which is maximally realized by the effect of the inserted central vowel, and the liquid. By dint of such comparison, Korean learners may easily identify the liquid after a consonant.

V. CONCLUSION

This study aimed to look at the impact of two main factors—visual salience and L1 background, on the use of visual cues in the perception of consonant contrasts by L2-learners of English. There was either mild or no AV benefit in the perception of non-native contrasts for L2 learners across learners with different L1s, and thus different relations of L2 contrasts to the phonological system in the L1. This therefore confirms the fact that the provision of visual cues for a second-language learner does not have the same universal 'enhancing' effect seen in native speakers where visual cues can help overcome the effects of hearing impairment, or environmental signal degradation. As predicted, the language background of the learner also impacted on the use of visual cues. Greater emphasis on auditory than visual cues was evident for Japanese-L1 speakers, and also quite markedly for Korean-L1 speakers, who achieved much higher scores for the /l/-/r/ contrast in the auditory than in the visual condition.

Significant correlations between scores in the auditory and visual conditions in this experiment, do suggest that increasing proficiency in the use of acoustic-phonetic cues to the non-native contrast leads to increasing proficiency in the use of visual cues to the contrast. This suggests that at least for visually-salient contrast, initial perceptual difficulties in the perception of the contrast using visual cues are due to the lack of appropriate phonemic labels for auditory cues rather than with difficulties in discriminating between these visual gestures.

REFERENCES

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*(4), 2299-2310.

Cole, R. A. (1999). Tools for research and education in speech science. *Proceedings of the International Conference of Phonetic Sciences,* San Francisco, CA.

Gordon, P. C., Keyes, L., & Yung, Y. F. (2001). Ability in perceiving nonnative contrasts: Performance on natural and synthetic speech stimuli. *Perception & Psychophysics, 63,* 746-758.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorisation in children aged 6 to 12. *Journal of Phonetics, 28,* 377-396.

Ingram, J. C. L., & Park, S.-G. (1998). Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America*, *103*(2), 1161-1174.

Im, B-B., & Ahn, H-S. (2002). Improving English listening comprehension by using animation. *English Language & Literature Teaching, 8*(2), 197-218.

Kang, H-S. (1999). Production and perception of English /r/ and /l/ by Korean learners of English: an experimental study. *Speech Sciences*, *6*, 7-24.

Kim, H.-J. (2004). Pronunciation error types and sentence intelligibility of Korean EFL learners. *English Language & Literature Teaching, 10*(3), 159-175.

Lee, B., & Lee, S-H. (2004). Korean learners' perception and production of English liquids. *Malsori*, *52*, 61-84.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*, Cambridge, MA: MIT Press.

Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contribution to speech perception. *Journal of Experimental Child Psychology, 41,* 93-113.

Ortega-Llebaria, M., Faulkner, A., & Hazan, V. (2001). Auditory-visual L2 speech perception: Effects of visual cues and acoustic-phonetic context for Spanish learners of English. *Proceedings of AVSP-2001*, 149-154.

Park, S.-G. (1999). /l/ and /r/ production by Korean and Japanese speakers of English: What factors are influential for the production? *Malsori, 37,* 87-117.

Sakow, M., & McNutt, J. (1993). Perception of /R/ by native speakers of Japanese and Korean: Internal and external perception. *The International Review of Applied*

Linguistics, 29(1), 46-53.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*, 212-215.

Summerfield, Q. (1983). Audio-visual speech perception, lipreading and artificial stimulation. In M. E. Lutman & M. P. Haggard (Eds.), *Hearing science and hearing disorders* (pp. 131-182), London: Academic Press.

Yamada, R. A. (1995). Age of acquisition of second language speech sounds: Perception of American English. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 305-320). Baltimore, MD: York.

Examples in: English
Application Languages: English
Applicable Levels: Elementary/Secondary/College/Higher

Hyunsong Chung
Dept. of English Education
Korea National University of Education
San 7, Darakri, Gangnaemyeon, Cheongwongun
Chungbuk 363-791, Korea
Tel: (043) 230-3554
Fax: (043) 232-7175
Email: hchung@knue.ac.kr

Received in April, 2005
Reviewed in May, 2005
Revised version received in June, 2005