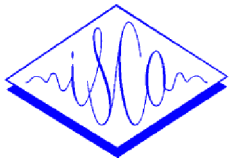
ISCA Archive
http://www.isca-speech.org/archive



Auditory-Visual Speech Processing
2007 (AVSP2007)
Hilvarenbeek, The Netherlands
August 31 - September 3, 2007

The impact of visual training on the perception and production of a non-native phonetic contrast

Valerie Hazan [1], *Anke Sennema* [2]

[1] Department of Phonetics and Linguistics, UCL, London, UK
[2] Institute of Linguistics, University of Potsdam, Germany
v.hazan@ucl.ac.uk, sennema@rz.uni-potsdam.de

Abstract

Many studies have shown that the perception of 'difficult' non-native phonetic contrasts can be improved through auditory training. More recently, studies comparing the effectiveness of auditory and audiovisual training have shown an advantage for audiovisual training at least for contrasts that are sufficiently visually-salient. Audiovisual training also led to improvements in the pronunciation of the trained consonants. The current study, which trained the /l/-/r/ contrast with Japanese learners of English, investigated training effectiveness using visual stimuli alone, i.e. with trainees seeing but not hearing the speakers. Fifteen Japanese students participated in the seven-session training programme and there were eleven controls. Pre/post tests were carried out in auditory (A), visual (V) and audiovisual (AV) test conditions and participants were also recorded before and after the training reading a list of words which included the sounds /l/ and /r/. Visual training was successful in significantly increasing the discriminability of the /l/-/r/ contrast in trainees in V and AV test conditions but there was no carry-over to the A condition. There was generalisation to nonsense words by unknown speakers. Visual influence was also tested by comparing performance in A and AV test conditions, and in a simple McGurk task. AV benefit increased to a greater extent for trainees. In the McGurk test, visual influence in the identification of discrepant (A /ba/- V /ga/) stimuli increased significantly in the post-test but this also occurred for controls, so might be due to a 'foreign-language' effect as most participants were attending a phonetics summer school in a foreign country. Results of an identification and rating test evaluating the participants' pronunciation of /l/ and /r/ pre- and post- training showed no evidence of any improvements in pronunciation following visual training.

Index Terms: Audiovisual perception, L2 speech perception, training, McGurk effect

1. Introduction

Second language (L2) learners often have difficulty in discriminating and pronouncing certain phonetic contrasts in the L2 that do not occur or have a different phonological status in their native language [e.g., 1]. However, even though there is a process of attunement to the sound categories of our native language in early childhood [e.g., 2], some plasticity remains. Programmes of targeted auditory training can be successful in increasing L2 learners' ability to discriminate such 'difficult' contrasts [e.g., 3], with good evidence of generalisation to new words and new speakers, and of long-term retention of the training effects [4].

Most training studies have only used auditory training materials. However, growing attention has also been focused on the perception of audiovisual speech in L2 speakers. It is generally accepted that speech is more intelligible when presented audiovisually both for native and nonnative speakers [e.g., 5]. However, analytic studies of phonetic categorization have found differences in the use of visual cues in L1 and L2 speakers. There is a debate as to whether such differences in the degree of visual influence in phonetic categorisation are language- or culture-specific, or indeed whether they exist at all. In a perception study comparing the auditory and audiovisual perception of difficult non-native contrasts, the amount of 'visual influence' in perception was shown to depend on the language background of the learner and on the degree of visual salience of the phonetic contrast [6]. In a study by the same researchers comparing auditory and audiovisual (AV) training of the /l/-/r/ and /b/-/v/ contrasts in Japanese learners of English [7], sensitivity to visual cues for L2 contrasts was enhanced following AV perceptual training. AV training was more effective in increasing the discriminability of the contrast than A training when the visual cues to the phonemic contrast were sufficiently salient. AV training, which enabled the learner to see the facial gestures of the speaker also led to a greater improvement in the learners' pronunciation of the contrasts, even for contrasts with relatively low visual salience. As with many studies of auditory training, this study reported individual variability in the effectiveness of training, and there is the possibility that AV training might not have been maximally effective for some learners, because of the greater cognitive load involved in attending to both the auditory and visual channels during training.

The aim of this study was therefore to extend these studies by assessing the effectiveness of visual training alone (i.e. with the trainees seeing but not hearing the speakers) for a contrast (/l/-/r/) with relatively low visual salience. It was hypothesised that this would focus the learners' attention on the differences in articulation between /l/ and /r/ (e.g., tongue movement, lip rounding). Another aim of this study was to see whether there would be any cross-modal effects: would training using lipreading alone lead to the trainees being able to better integrate the auditory and visual information in audiovisual conditions? Following increased attention to articulatory differences between /l/ and /r/, would there be any cross-modal effect in an auditory alone condition, as might be suggested by models of speech perception that imply strong links between perception and production? Finally, are there any significant correlations between the perception and production of the /l/-/r/ contrast prior to training and does visual training result in any changes in production?

2. Visual training

2.1. Participants and recording procedure

Twenty-six Japanese-L1 learners of English as a Foreign Language participated in the study: 15 trainees and 11 controls. All of the trainees and most of the controls were resident in Japan and attending a two-week English phonetics summer school in London. Five of the controls were Japanese participants who were resident in the UK. Learners were approximately at a lower to lower-intermediate level of English proficiency, aged between 19 and 40 years (median: 20.5 years) and had typically started learning English at school at the age of 12, but with a focus on written language. The majority had never lived in an English-speaking country. A further six students participated in the study but were excluded because their pre-test performance in the nonsense word test was above 90% in the AV condition.

Two male and one female speaker of South Eastern British English recorded the items for the pre/post-tests, and two further women and three men recorded items for the training materials and a generalisation test which is not reported here. Video recordings were made in a soundproof room, with the speaker's head fully visible within the frame. The video and audio channels were digitally transferred to a PC. Video clips were edited so that the start and end frames of each token showed a neutral facial expression. Stimuli were down-sampled post-editing (250*300 pixels, 25 f/s, audio sampling rate 22.05 kHz).

2.2. Materials

2.2.1 Training materials

For the training sessions, materials included 71 minimal pairs of real words: 38 containing /l/ or /r/ as singletons in initial position and 33 containing /l/ or /r/ in initial clusters.

2.2.2. Pre/post nonsense word test

The two consonants /l/ and /r/ were embedded in nonsense words in initial (CV, cCV) and intervocalic (VCV) position in the context of the vowels /i, a, u/. In the cCV stimuli, the consonant clusters were /pl/, /br/, /gl/, /cr/. The test included two repetitions of each item produced by each of three speakers for initial singletons (total: 36 stimuli), and one repetition for initial clusters (total: 36 stimuli) and medial singletons (total: 18 stimuli). The 90 items were randomised and presented in a single block with a pause after 50 trials.

2.2.3. Pre/post McGurk test

To further test visual influence and audiovisual integration, a short McGurk test was included in the pre/post test battery. The stimuli for this test were taken from materials used in a study on the McGurk effect in Taiwanese and English children and adults [8]. Items were recorded by two English and two Taiwanese speakers (one man and one woman for each language group). For each speaker, items in the AV condition included four repetitions of one congruent /ba/ item (auditory-visual /ba/), of one congruent /ga/ item (auditory-visual /ga/) and of one incongruent item (auditory /ba/ combined with visual /ga/). In the Auditory (A) condition, for each speaker, materials included four repetitions of /ba/ and of /ga/. In the Visual (V) condition, for each speaker, materials included three repetitions of the syllables /ba/, /da/ and /ga/. Full details of the stimuli can be found in [8].

2.3. Experimental task

Test and training programmes, designed using the CSLU toolkit [9], ran individually on desktop computers. Students worked in quiet surroundings in the presence of an experimenter, and stimuli were presented via headphones at a comfortable listening level. In the pre-training test session, all participants carried out the nonsense-word test first, then the McGurk test. The post-training test was identical to the pre-test for controls but the trainees additionally carried out a generalisation test. The pre/post tests took around 40 minutes to complete.

In the pre/post tests, the nonsense-word tests were presented in three conditions: 'auditory' (A), 'audiovisual' (AV), and 'visual' (V) in a two-alternative forced-choice identification task, with response choices of R and L, and no feedback given. Two orders of presentation (AV, A, V or A, AV, V) were counterbalanced across participants. Participants responded by clicking on the appropriate letter symbol with a mouse. The test items were identical in all conditions, but in the single-modality tasks, either the auditory or visual channels were removed. Each block of 90 items was presented once in each test condition, yielding a total of 270 responses per participant. The McGurk tests were also presented in three conditions in the same order as the nonsense word test for that participant.

For the trainees, the pre-test was followed by seven sessions of visual training, each lasting about 40 minutes, carried out within a two- week period. In these sessions, trainees saw video clips of the speakers producing words with initial /l/ or /r/ but but did not hear them. The program was run individually on desktop PCs, with trainees working in quiet surroundings under the supervision of an experimenter. After each presentation, the trainee had to click on L or R on the screen. If the response was correct, a 'smiley' appeared. If the response was incorrect, a sad face appeared and the video was presented again with the correct label shown. At the end of each block, a bar chart showed the percentage of correct R and L responses, with a message of encouragement or congratulations. At each training session, listeners first heard two blocks of test items produced by one speaker: 76 'singleton' tokens and 66 'cluster' tokens. After a short pause, a second speaker was presented with again a 'singleton' block followed by a 'cluster' block. Over the seven days of training, trainees saw four of the speakers three times and one of the speakers twice.

Control participants carried out the pre- and post-test only, separated by the same amount of time as the trainees. The majority of control participants were also Summer School students, so were receiving some English phonetics tuition over the two-week period, but no specific visual training.

2.4. Results

2.4.1. Pre/post nonsense word tests

The percentage of correct consonant identification obtained in each condition was calculated. In the pre-test, over all participants (N=26), mean /l/-/r/ identification was 63.4%, (s.d. 13.4) in the A condition, 64.1% (s.d. 12.8) in the AV condition and 57.5% (s.d. 11.2) in the V condition. In order to

correct for any potential bias in responses, scores were converted to the signal detectability measure dprime (d'), calculated as the z-value of the hit-rate minus that of the false-alarm-rate. First, the pre-test results were examined to compare participants' performance in the different modalities (See Table 1). A repeated-measures ANOVA showed that there was no evidence of 'AV benefit' (AV>A) in either the trainees or controls.

Table 1. *Means and standard deviations for the score (d') representing the discriminability of the /l/-/r/ contrast in the pre- and post-test in each test condition (A, AV, V) for the training and control groups.*

Group	Cond	Pre-test	Post-test	Diff post-pre
Training (N=15)	A	0.56 *(0.56)*	0.84 *(0.66)*	+0.28 *(0.51)*
	AV	0.59 *(0.54)*	1.51 *(0.58)*	+0.93 *(0.60)*
	V	0.18 *(0.54)*	1.05 *(0.38)*	+0.87 *(0.60)*
Controls (N=11)	A	1.05 *(1.06)*	1.3 *(1.16)*	+0.25 *(0.51)*
	AV	1.04 *(0.90)*	1.43 *(1.08)*	+0.39 *(0.38)*
	V	0.68 *(0.58)*	0.68 *(0.55)*	-0.00 *(0.30)*

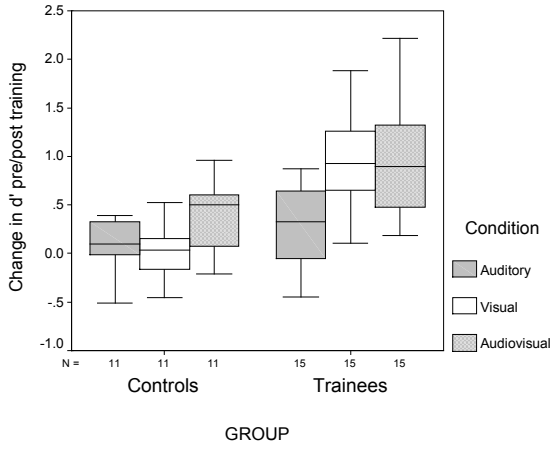


Figure 1: *Boxplot showing Change in discriminability (d') of /l/-/r/ contrast between the pre-test and post-test for trainees and controls.*

The evaluation of training effectiveness was carried out on the difference between pre-test and post-test d' scores so as not to be influenced by any differences in baseline performance levels (See Figure 1). Trainees showed a greater increase in performance than controls [$F(1, 24)= 14.48$; $p<0.001$]. The effect of test condition was significant [$F(2, 48)= 5.25$; $p<0.01$], and there was a significant test condition by group interaction [$F(2, 48)= 5.25$; $p<0.01$]: both groups showed similar increases in discriminability for the A condition but trainees improved more in both the V and AV conditions.

The data was then specifically examined to see whether there was evidence of an increase in AV benefit (AV>A) as a result of training. For the control group, there was no difference between the A and AV conditions in either the pre- or post-test. For the training group, there was no difference in discriminability across the A and AV conditions pre-test, but there was evidence of AV benefit post-training as discriminability was significantly higher in the AV condition than in the A condition. There was also a significant difference between performance in the AV and V conditions.

2.4.2. Pre/post McGurk tests

As the pre-test data was missing for two participants, the data analysis was carried out on data from 14 trainees and 10 controls. First the AV condition was examined. As few repetitions were collected per item, data was grouped across the four speakers to get scores representing 'auditory congruent' responses to the AV-congruent /ba/ and /ga/ items and to the discrepant 'A /ba/-V /ga/' items, which were most likely to show the degree of visual influence. The responses to the congruent stimuli were as expected with no responses in the post-test being lower than 100% correct and only four in the pre-test. Results for the discrepant items are presented in Figure 2. It must be noted that the overall rate of auditory-congruent responses was low for the Japanese speakers, and therefore that they were showing a strong McGurk effect. A repeated-measures ANOVA showed that there was a decrease in A-congruent responses (i.e., an increase in visual influence) in the post-test relative to the pre-test [$F(1, 22)= 9.918$; $p=0.005$] but this was the case for both trainees and controls.

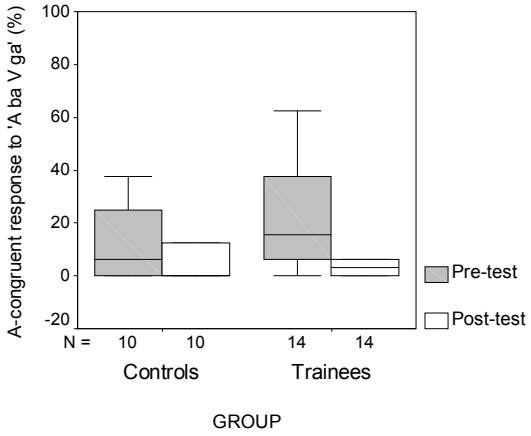


Figure 2: *Rate of auditorily-congruent responses for the (/ga/ audio-/g/ visual)items in the McGurk AV test.*

For the V condition, mean scores for the trainees were 70.2% (s.d. 10.8) in the pre-test and 67.6% (s.d. 8.6) in the post-test; mean scores for the control group were 66.4% (s.d. 8.1) in the pre-test to 68.6% (s.d. 10.4) in the post-test. Repeated-measures ANOVA confirmed that there was no significant effect of group or time of testing. The training group did therefore not improve in lipreading the BA-DA-GA contrast (McGurk test) after visual training with /l/-/r/ stimuli.

2.5. Discussion

This study shows that purely visual training of the /r/-/l/ contrast was successful in increasing the discriminability of this contrast, despite its relatively low visual salience. These increases were shown not only in the V test condition but also

in the AV condition, with evidence of AV benefit in the post-test only for the trainee group. It should be noted that this visual training effect was obtained with Japanese speakers who are generally thought to be relatively insensitive to phonetic visual information in Japanese [10], although they usually show greater visual influence with non-native speakers. Here, all speakers were non-native for the trainees. As was the case for the AV training carried out in our previous study [7], the impact of training modality can be seen on the specific channel trained and in greater AV integration, but with minimal impact on the channel that was not trained. There seems to be little evidence of cross-modal effects of training. Information that was learned about the articulatory gestures characteristic of /l/ and /r/ did not assist the listener when decoding acoustic cues to these contrasts.

An increase in influence of visual cues was also shown in the McGurk test, although one needs to be cautious about these results due to the relatively small amount of data collected per listener. The first surprising finding is that the general level of McGurk effect overall was much higher than reported in some previous studies with Japanese speakers [e.g., 12]. A high level of McGurk responses for these same stimuli was also obtained with English, Taiwanese, Thai and Japanese adults in a separate study [13]. As it is quite typical to use a relatively small number of discrepant items in McGurk studies (with often the same items repeatedly used in different studies by the same research group), the degree of McGurk effect may be partly dependent on the particular items used. This high level of visual influence may also be due to the fact that all four speakers used in the test were 'foreign' to the Japanese listeners (two English, two Taiwanese). Indeed, a stronger McGurk effect has been reported in Japanese listeners when listening to foreign speech even though this did not reach the rates shown here [12]. Another explanation is that visual salience might have increased in both trainees and controls over the two week period because they were attending a demanding summer school in phonetics in a foreign country. However, there was little relation between improvements seen in /l/-/r/ in the V condition following visual training and results in the V condition of the BA-DA-GA contrast. This suggests that improvements in the /l/-/r/ V condition following training are linked to the learning of specific visual information about the /r/-/l/ contrast rather than merely due to an increased focus on to visual cues.

3. Production study

The second aim of the study was to investigate if purely visual phonetic training could result in an improvement in the pronunciation of the difficult /l/-/r/ contrast. A significant improvement in speech production had been obtained in our previous study after audiovisual training of the same contrast [7].

3.1. Participants

Ten native English speakers participated in the production rating study.

3.2 Materials

Each participant in the training study (trainees and controls) recorded a list of twelve words with initial /l/ and /r/ randomised three times at the end of both the pre-test and post-test sessions. Recordings were made in a sound-treated room at a sampling rate of 22050 Hz. Four of the minimal pairs (lake-rake, long-wrong, blight-bright, clash-crash) were used in the identification and rating test; the second repetition of each word was edited out and saved to individual files. An example of productions of 'lake' and 'rake' by one of the trainees prior to training (lake_rake_pre.wav) and following training (lake_rake_post.wav) is included.

3.3 Experimental task

An identification and rating task was prepared using PRAAT (version 4.6.01). In the test, participants heard a word and had to first decide whether the word contained a R or a L by clicking on the appropriate label, and then had to give a rating from 1 (bad) to 7 (excellent) in terms of how the R or L sounded. Participants could replay the item up to three times if needed. Each native listener identified and rated 4 tokens per consonant produced at pre- and post-test sessions for each Japanese speaker, so 4160 ratings were obtained across the ten native listeners (26 speakers * 4 tokens * 2 consonants * 2 times).

3.4. Results

For one speaker, one /l/ token and one /r/ token from the pre-test were missing. They were replaced by duplicating the scores achieved for these tokens in the post-training, thus assuming no change.

The means of correct consonant identification were calculated for each of the 15 trainees and 11 controls (see Table 2).

	/l/		/r/	
	pre	post	pre	post
Trainees				
Mean (%)	59.67	64.33	78.17	84.67
S.D.	26.99	25.80	29.50	27.34
Controls				
Mean (%)	70.90	72.73	86.13	88.18
S.D.	20.13	16.33	29.77	15.33

Table 2: *Means and standard deviations for the percentage of correct consonant identification*

The identification of /l/-tokens was in general poorer than the identification of /r/-tokens which is consistent with previous studies (e.g. [6], [14]). Tokens produced by the control group achieved a higher rate of correct identification in the pre-test than those produced by the training group. This mirrors their higher perception scores in the pre-test as well. From pre-test to post-test scores, mean identification scores increased by 4.7 % for /l/, and 6.6 % for /r/ for the trainee group but only by 1.83% for /l/ and 2.05% for the control group. However, a univariate ANOVA carried out on pre/post-test data showed that there was no significant difference in both groups between the pre and post-test identification scores.

The English native listeners also judged the quality of /l/ and /r/ realizations in each token. For each of the Japanese speakers, mean consonant rating scores were calculated for /l/ and /r/ only for those items with correct identification (see Table 3).

As for the identification results, there were no statistically significant differences between the ratings obtained for the correctly-identified pre- and post-test tokens for either the trainees or controls.

	/l/		/r/	
	pre	post	pre	post
Trainees				
Mean	4.80	4.47	5.05	4.91
S.D.	1.75	1.86	1.39	1.50
Controls				
Mean	4.76	4.64	5.09	4.87
S.D.	1.55	1.61	1.37	1.54

Table 3: *Consonant accuracy rating on a scale of 1-7 (1=bad, 7=excellent), by 10 native English listeners.*

3.5 Discussion

In our previous study, we found that AV trainees showed a greater improvement in their pronunciation of the /l/-/r/ contrast than A trainees, even though the AV trainees had not improved their perception of the contrast to a greater extent than trainees tested with auditory stimuli [7]. We suggested then that exposure to the articulatory gestures involved in the production of /l/ and /r/ seemed to have been effective in improving pronunciation, even without specific pronunciation training. Here, trainees were also exposed to this articulatory information, with even greater emphasis placed on visual cues in the absence of acoustic information. However, mere exposure to visible articulations did not appear to have the same effect on production than combined exposure to visible articulations and to acoustic information about the contrast. As different speakers were used in the pre/post tests in these two studies and as AV trainees also had 3 more training sessions, further investigation is needed to confirm this finding.

4. Relation between perception and production results

Spearman's correlations were used to look at significant relations between production (i.e., the intelligibility of participants' productions of /l/ and /r/) and perception (performance on the nonsense word and McGurk tests). First, the data was examined for the group as a whole (N=26) for the pre-test results to look at correlations prior to training. The production score was significantly correlated with perception scores on the AV pre-test (ρ=.473(24), p<.02) and A pre-test (ρ=.580(24), p<.005) but not with the V pre-test or McGurk scores (V condition or 'discongruent' AV stimuli). Closer examination revealed that these correlations were only valid for the control data; in the training group, there were no significant correlation between production and perception measures.

In the post-test, over the whole group (N=26), post-test production was again correlated with the nonsense word perception test in the A (ρ=.742(24), p<.001) and AV conditions (ρ=.596(24), p<.001) but not the V condition or McGurk scores. For the trainees alone, post-training production was only significantly correlated with A post-test (ρ=.632(13), p<.02).

In summary, the ability to accurately pronounce /l/ and /r/ seems to be primarily correlated with the ability to hear the contrast accurately, but not with the ability to discriminate these sounds visually.

Conclusions

These results suggest that visual training alone is successful in increasing learners' use of visual information to this difficult phonetic contrast, even though the visual contrast between /l/ and /r/ is not highly salient even to native listeners. However, although this training led to an improvement in perception in the V and AV conditions, there was no cross-modal effect on perception using the auditory channel alone. Also, increases in visual influence were not linked to an improvement in the production of these difficult sounds. Visual training seems therefore less successful than AV training, which led to improvements in pronunciation. It must be noted that our two studies (of A and AV training in previous study and V in current) cannot be directly compared because the current study involved different speakers in the pre/post tests and different numbers of training sessions. A controlled study with common training and test materials but with training in different modalities is therefore necessary to confirm current findings. Also, given the greater success of the visual training in increasing visual influence, a 'hybrid' training programme initially involving single modalities (auditory and visual training) and then moving on to combined audiovisual training may be the most effective option.

References

[1] Flege, J.E., "Second-language speech learning: theory, findings, and problems". In: Strange, W. (ed.), Speech Perception and Linguistic Experience: Theoretical and Methodological Issues. Baltimore: York Press, 229-273, 1995.

[2] Kuhl, P. K., "Early language acquisition: Cracking the speech code", Nature Reviews Neuroscience, vol. 5, 2004, pp. 831-843.

[3] Logan, J.S., Lively, S.E., Pisoni, D.B., "Training Japanese listeners to identify English /r/ and /l/: A first report", J. Acoust. Soc. Am., vol. 89, 1991, pp.874 - 886.

[4] Lively, S. E., Logan, J. S., Pisoni, D.B., "Training Japanese listeners to identify English /r/ and /l/. III: long-term retention of new phonetic categories", J. Acoust. Soc. Am., vol. 96, 1994, pp. 2076-2087.

[5] Hardison, D., "Variability in bimodal spoken language processing by native and nonnative speakers of English: A closer look at effects of speech style", Speech Communication, vol. 46, 2005, pp. 73-93.

[6] Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., Chung, H., "The use of visual cues in the perception of nonnative consonant contrasts", J. Acoust. Soc. Am., vol. 119, 2006, pp. 1740-1751.

[7] Hazan, V., Sennema, A., Iba, M., Faulkner, A., "Effect of audiovisual perceptual training on the perception and production of consonants in Japanese learners of English", Speech Communication, vol. 47, 2005, pp. 360-378.

[8] Chen, Y. and Hazan, V., "Developmental Factor in Auditory-Visual Speech Perception-The McGurk Effect in Mandarin-Chinese and English Speakers", Proceedings of AVSP2007, Netherlands, 1-3 September 2007.

[9] Cole, R., Massaro, D.W., de Villiers, J., Rundle, B., Shobaki, K., Wouters, J., Cohen, M., Beskow, J., Stone, P., Connors, P., Tarachow, A., Solcher, D., "New tools for interactive speech and language training: using animated conversational agents in the classroom of

profoundly deaf children", Proc. ITRW on Methods and Tools in Speech Science Education (MATISSE), London, 1999, pp. 45-52.

[10] Sekiyama, K. and Tohkura, Y., "Inter-language differences in the influence of visual cues in speech perception", J. Phonetics, vol. 21, 1993, pp.427-444

[11] Sekiyama, K. (1997) Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. Perception & Psychophysics, 59, 73-80.

[12] Sekiyama, K., Burnham, D., Tam, H. and Erdener, D., "Auditory-Visual Speech Perception Development in Japanese and English Speakers", In Proceedings of the International Conference on Auditory-Visual Speech Processing, St. Jorioz, France, 2003, p61-66

[13] Chen, Y. and Hazan, V., "Language effects on the degree of visual influence in audiovisual speech perception", Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrueken, Germany, 6-10 August 2007.

[14] Bradlow, A., Pisoni, D., Akahane-Yamada, R. and Tohkura, Y., 1997. Training Japanese listeners to identify English /r/ and /l/: IV. some effects of perceptual learning on speech production, J. Acoust. Soc. Am., vol. 101, 1997, pp. 2299-2310.