Inceoglu, S. (2015). Audiovisual and auditory-only perceptual training: Effects on the pronunciation of French nasal vowels. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds). *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (ISSN 2380-9566), Santa Barbara, CA (pp. 104-114). Ames, IA: Iowa State University.

AUDIOVISUAL AND AUDITORY-ONLY PERCEPTUAL TRAINING: EFFECTS ON THE PRONUNCIATION OF FRENCH NASAL VOWELS

Solène Inceoglu, Rochester Institute of Technology

Previous studies have shown that gains made during perceptual training can be transferred to gains in production and that audiovisual (AV) perceptual training often leads to greater improvement in production than auditory-only (A-only) training (Hardison, 2003; Hazan, Sennema, Iba, & Faulkner, 2005). This study investigated whether perceptual training on the three French nasal vowels led to improvement in the production of these vowels, and whether improvement was greater with AV perceptual training as opposed to A-only training. The productions of 60 American-English intermediate learners of French were recorded at pretest and posttest. The stimuli consisted of 108 CVC words in various consonantal contexts and the L2 learners' productions were judged by two native French listeners in two rating tasks: a forced-choice identification rating task and a quality rating task. Results showed that both training groups—but not the control group— significantly improved from the pretest to the posttest, but that the production of the AV training group improved significantly more than the production of the A-only training group. Furthermore, the two types of analysis used to assess the production of the L2 learners revealed differences that have implications for research methodology and assessment.

INTRODUCTION

There is a general consensus that most individuals who learn a second language (L2) as adults speak it with a foreign accent (e.g., Derwing & Munro, 1997; Flege, Munro, & MacKay, 1995; Major, 2001). This fact has led to proposals for several model of speech perception. The Speech Learning Model (Flege, 1992, 1995), for instance, claims that the similarity between L1 and L2 sounds makes it difficult for learners to perceive phonetic differences and to create new categories for L2 sounds. The Similarity Differential Rate Hypothesis (Major & Kim, 1996) goes further and argues that dissimilar sounds between an L1 and L2 are acquired faster than similar sounds. This crosslinguistic influence in perception is also believed to be reflected in production, and numerous studies have shown that the degrees of accuracy in perceiving and producing L2 phones are related (e.g., Flege, Bohn, & Jang, 1997; Flege, 1988). Of interest to second language acquisition (SLA) researchers and teachers is the question of how novel L2 phoneme categories are acquired and the extent to which instruction or training can improve production. A wealth of empirical studies have provided evidence that perceptual auditory training can be transferred to improvement in production even when no production tasks are involved during training (e.g., Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Lopez-Soto & Kewley-Port, 2009; Wang, Jongman, & Sereno, 2003).

A number of different methods have been used assess L2 learners' speech production. Among the tasks commonly used are forced-choice identification tasks (Bradlow et al., 1997; Hazan et al., 2005), identification tasks (Lambacher et al., 2005), quality rating tasks (Hardison, 2003;

Hazan et al., 2005), acoustic analyses (Lambacher et al., 2005; Wang et al., 2003) and pretest-posttest paired comparison (Bradlow et al., 1997). In addition, while some studies use only one type of rating task (e.g., Hardison, 2003; Lopez-Soto & Kewley-Port, 2009; Wang et al., 2003), others combine several tasks (e.g., Bradlow et al., 1997; Hazan et al., 2005; Lambacher et al., 2005).

Audiovisual Speech Perception

The large majority of L2 speech studies have focused on one source of input—the auditory signal—despite the fact that speakers rely on both auditory and visual information in face-to-face communication, making communication a multimodal experience (Pisoni & Remez, 2005; Rosenblum, 2005). The ability to benefit from both the auditory and visual modality in perceiving and interpreting speech has been demonstrated with native speakers in studies investigating L1 speech comprehension (e.g., Arnold & Hill, 2001; Reisberg, McLean, & Goldfield, 1987; Summerfield, 1979), speech intelligibility (e.g., Benoît, Mohamadi, & Kandel, 1994; Sumby & Pollack, 1954), and language discrimination (e.g., Ronquest, Levi, & Pisoni, 2010; Soto-Faraco et al., 2007). For instance, Arnold and Hill (2001) found that comprehension was better in audiovisual conditions than in auditory-only conditions when messages where presented in various ways: with accented speech, in a L2 that participants were fluent in, or with complex semantic and syntactic structures. In addition, although non-native speakers have been found to be less efficient at using visual information than native speakers (e.g., Hazan et al., 2005, 2006; Ortega-Llebaria et al., 2001), the superiority of audiovisual information over auditory-only information has also been supported in studies looking at L2 speech perception (e.g., Erdener & Burnham, 2005; Hardison, 2003; Hazan et al., 2006, 2005; Hirata & Kelly, 2010; Kluge, Reis, Nobre-Oliveira, & Bettoni-Techio, 2009; Wang, Behne, & Jiang, 2008).

Despite a growing number of studies pointing to the facilitative aspects of audiovisual speech information in attending to L2 speech, only a couple of studies have examined the relationship between audiovisual speech perception and production (Erdener & Burnham, 2005) or looked at the effect of audiovisual training on the production of L2 phonemes (Hardison, 2003; Hazan et al., 2005). Because of the relative paucity of published studies, it is premature to draw firm conclusions. Nonetheless, two studies investigating the production of the English /r-l/ contrast by Japanese speakers reported that audiovisual perceptual training led to greater improvement in production than auditory-only training (Hazan et al., 2005). Production performance was also found to be superior when visual information (i.e., the face of the speaker) was present in a study where Australian English and Turkish speakers were asked to repeat words produced in Spanish and Irish (Erdener & Burnham, 2005).

Still, much work remains to be done regarding the effect of L2 speech perception on speech production. Accordingly, this study contributes to the ongoing research on audiovisual speech perception by exploring whether perceptual training helps improving the production of the three (Standard) French nasal vowels. The rationale for targeting nasal vowels is that their difference is visually salient, ranging from the hyper-rounded [ɔ̃] to the rounded [ɑ̃] and unrounded [ɛ̃], and they are often problematic for non-native speakers.

Research Questions

The general research question addressed in this article was whether participants receiving audiovisual (AV) perceptual training improve their pronunciation of the French nasal vowels more than participants receiving audio-only (A-only) training. In order to have a better understanding of assessment methods and rating tasks, the research question was divided into two sub-questions examining participants' pronunciation when (a) rated by native speakers in a quality rating task, and (b) rated by native speakers in a forced-choice identification task.

METHODS

Participants

Sixty participants (age 18-24, mean = 20; 43 females) were recruited from intermediate French language courses at a large Midwestern university and were randomly assigned to one of the three following groups: AV training, A-only training, and a control group. All the participants were native speakers of American English who reported good vision, no hearing disorder, and no background in lipreading and phonetics.

Stimuli and Procedure

The participants were tested at pretest and posttest individually in a quiet room. A delayed repetition task was used to elicit participants' production of 108 monosyllabic #CVC# words where the vowel was either [ɑ̃], [ɔ̃], or [ɛ̃]. The initial consonant was one of the following: [p-t-k-b-d-g-s-z-f-v-ʒ-ʃ] to take into consideration the articulation of vowels in different consonantal contexts. Participants heard the stimuli followed by one of five version of a prompt asking them, in French, to repeat the word. For example, they would hear the stimulus "ponse" [pɔ̃s] followed by "*répète le mot s'il te plaît*" (repeat the word, please). The reason for having a prompt between the stimuli and the participants' repetition of the stimuli was to prevent direct imitation from sensory memory.

Two rating tasks were used to assess participants' oral production. Two native listeners of French rated the pretest and posttest productions of the sixty participants in a forced-choice identification task and a quality rating task. For both tasks, raters were asked to focus on the accuracy of the vowels production and ignore the production of the consonants. In the forced-choice identification task, the raters heard a stimulus produced by L2 participants and were asked to choose which of the three nasal vowel the L2 participants had produced. The task was not timed, and raters had the option to listen to the production a second time, if needed. The two native listeners rated 12,960 tokens (108 stimuli × 2 tests (pre and post) × 60 participants) and the inter-rater reliability was 88.4%, which is considered acceptable. Discrepancies were rated by a third native rater to ensure that all responses used for the analysis were agreed on by at least two native speakers of French. In the quality rating task, raters were presented with a target word on a computer screen, listened to a participant's production of the word and rated the production on a scale from 1 (bad) to 7 (excellent). Participants' pretest and posttest productions of each single word were presented in one block, resulting in a total of 108 blocks—one for each word. The average between the ratings of the two native listeners was calculated and used for the

quality rating task. The order of the production tokens was randomized across raters and the ratings were completed within a period of one month.

Perceptual Training

Between the production pre- and post- tests described above, the two training groups received six 30-minute sessions of either AV (hearing a speaker while seeing her face) or A-only (just hearing the speaker) high-variability phonetic perceptual training. Participants sat in front of a computer screen and were presented with a #CVC# stimulus similar to those used for the production task. They then had 4000 milliseconds to click on one of the three options (e.g., nasal vowel "an", "on", or "un") on the screen before receiving feedback. Participants were presented with 178 CVC stimuli randomized across participants and training sessions.

RESULTS

Forced-choice Identification Rating Task

The mean production accuracy scores at pretest and posttest are illustrated in Figure 1. A repeated-measures ANOVA with Time as the repeated measure and Group as between-subject factor was conducted to measure the effect of perceptual training on production accuracy. Results indicated a significant effect of Time, $F(1, 5989) = 111.17$, $p < .001$, as well as a significant interaction between Group and Time, $F(2, 5989) = 24.16$, $p < .001$. This interaction was analyzed with separate t-tests (adjusted alpha level of .016) which showed that there were significant differences between the pretest scores (M = .63, SD = .48) and posttest scores (M = .76, SD = .42) of the AV group, $t(1833) = 10.20$, $p < .001$, as well as between the pretest scores (M = .66, SD = .47) and posttest scores (M = .73, SD = .44) of the A-only group, $t(2110) = 6.38$, $p < .001$. There was, however, no significant differences between the pretest scores (M = .68, SD = .46) and posttest scores (M = .70, SD = .46) of the control group, $t(2046) = 1.14$, $p = .254$, providing evidence that improvement in production accuracy was related to perceptual training. A one-way ANOVA with Training as the between-subjects factor and Gain change as the dependent variable revealed that the improvement in production accuracy of the AV training group was significantly greater than the improvement of the A-only training group, $F(1, 3944) = 11.95$, $p = .001$, thus confirming the superiority of AV perceptual training over A-only training as far as transfer to production is concerned.

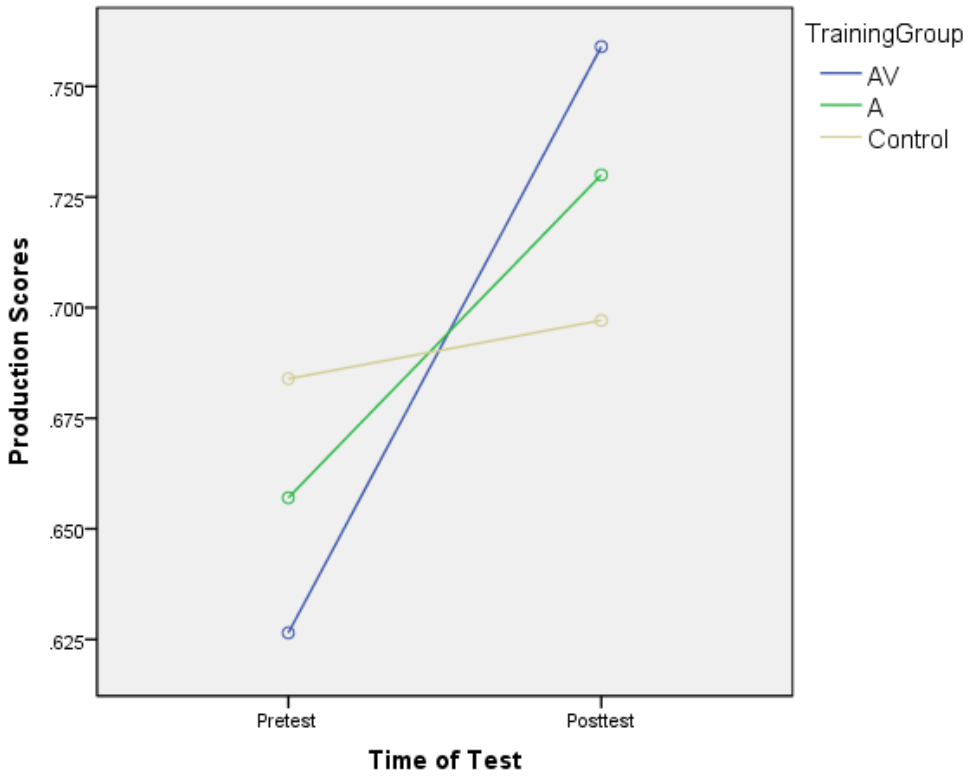

Figure 1. Accurate production at pretest and posttest as rated by native speakers in a forced-choice identification task (maximum score 1).

Further analyses were conducted to compare production in each of the three vowels (Table 1). Results indicate that AV training led to significantly greater improvement in the pronunciation of [ɔ̃] than A-only training (p = .014) and that the two types of training led to greater improvement than no training. The difference in the improvement of [ɑ̃] was not statistically different between the two training groups (p = .59) or between the A-only training and control groups (p = .225), but the AV training group improved more than the control group (p = .028). Finally for [ɛ̃], the AV training group improved significantly more than the A-only training (p = .046) and the control group (p < .001), but the A-only training did not improve more than the control group (p = .185).

Table 1

Percentage of correct identification by native speakers of vowels produced by L2 speakers at pretest and posttest

		[ɔ̃]	[ɑ̃]	[ɛ̃]
AV training	Pretest	58.75	54.99	74.18
	Posttest	77.22	58.83	90.12
A-only training	Pretest	60.77	56.18	78.88
	Posttest	72.01	58.23	88.79
Control	Pretest	63.54	60.96	80.79
	Posttest	65.64	56.72	86.80

Quality Rating Task

Results of the quality rating tasks are illustrated in Figure 2. Repeated-measures ANOVAS showed that all groups significantly improved: $F_{AV}(1, 1833) = 152.73$, $p < .001$; $F_{A}(1, 2157) = 43.00$, $p < .001$; $F_{C}(1, 2048) = 33.91$, $p < .001$. A one-way ANOVA with Group as the between-subjects factor and Gain change as the dependent variable revealed a significant effect of Group, $F(2, 6040) = 7.70$, $p < .001$ and a Bonferroni post-hoc test identified that the AV group improved significantly more than the A-only group ($p = .002$) and the control group ($p = .002$). However, and contrary to the overall results of the forced-choice identification task, the A-only group did not appear to have improved more than the control group did ($p = .10$).

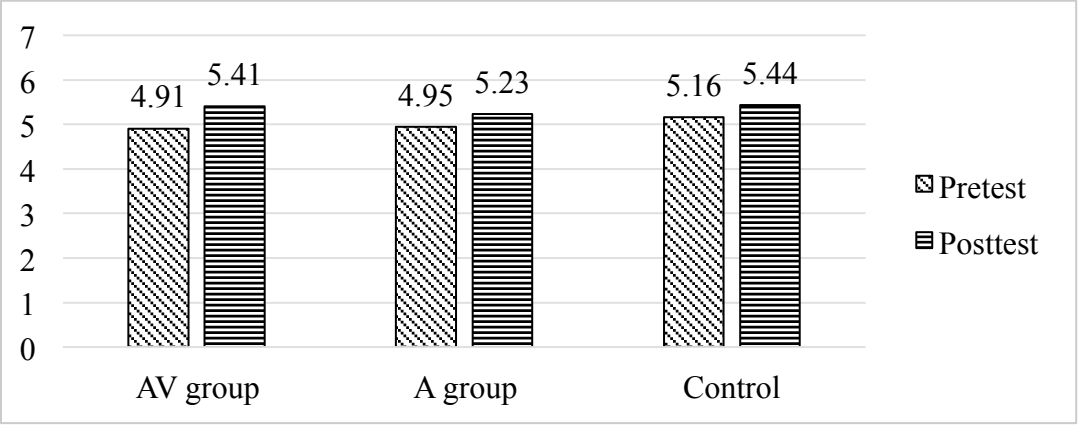

Figure 2. Percentage of accurate production at pretest and posttest as rated by native speakers in a quality rating task**.**

Detailed analyses for each vowel (Table 2) showed that the difference in the improvement of the pronunciation of [ɔ̃] by the two training groups was not significant ($p = .094$), and that the two training groups improved significantly more than the control group (AV: $p < .001$, A: $p = .029$). As for [ɑ̃], not only did the AV training group improve significantly more than the A-only training group ($p = .022$), but the performance of the latter actually decreased

from pretest to posttest. Furthermore, the difference in pronunciation improvement between the control group and the two other groups did not reach significance (AV: $p = .352$, A: $p = .405$). Finally, there was no significant effect of Group, $F(2, 2013) = 0.70$, $p = .494$ for the improvement of [ɛ̃], indicating that all groups improved in a similar way.

Table 2

Means of production rating (7-point scale) for each vowel at the pretest and posttest

		[ɔ̃]	[ɑ̃]	[ɛ̃]
AV training	Pretest	4.98	4.69	5.07
	Posttest	5.62	4.9	5.7
A-only training	Pretest	4.94	4.82	5.09
	Posttest	5.39	4.74	5.58
Control	Pretest	5.29	4.85	5.35
	Posttest	5.49	4.91	5.93

DISCUSSION

The primary goal of this study was to compare the effect of audiovisual and audio-only perceptual training on the production of French nasal vowels by intermediate American-English learners of French. Results showed that, overall, after six sessions of perceptual training, the pronunciation of the learners in the two training groups improved significantly more than the pronunciation of the control group. These results are comparable to the perceptual training studies mentioned earlier and demonstrate that improvement in pronunciation can be achieved using a simple perceptual training task. More importantly, the results of this study demonstrated that AV perceptual training was more efficient than auditory-only training, and that participants in the AV training group benefited from the visual information they received during training and that they transferred their knowledge onto production.

The results also showed that the effects of perceptual training were stronger for the vowel [ɛ̃], which was about 90% accurately produced by both training groups at posttest. Perceptual training did not, however, seem to be as beneficial for [ɑ̃], which showed little improvement from pretest to posttest. This may be due to the lack of visual saliency of that vowel, which is neither hyperrounded, like [ɔ̃] nor unrounded like [ɛ̃].

The secondary goal of this study was to compare two types of rating tasks to get a better understanding of the data. The two tasks used showed a general similar pattern, namely that AV training was superior to A-only training, but also revealed some differences. It has previously been noted that differences in results across studies may be due to the use of different testing materials and rating procedures (Flege, MacKay, & Meador, 1999), settings, language background, and language proficiency. In the current case, the differences in results can only be explained by the rating tasks used since the participants and raters remained the same. The forced-choice identification task provided a stronger base for the argument in favor of AV training: for all vowels, the AV trainees improved significantly more than the control group, but

the AV training did not seem to be more effective than auditory-only training only for the vowel [ɑ̃], probably again because of the vowel's lack of visual saliency. On the other hand, the quality rating task led to a less straightforward picture. Results indicated that in terms of vowel quality, the AV trainees' pronunciation of [ɑ̃] improved significantly more than the pronunciation of the A-only trainees, but that AV training did not help to improve the pronunciation of [ɔ̃] and [ɛ̃] more than A-only training did. Finally, results of the quality task revealed that no group was better than the other two groups at improving the pronunciation of the vowel [ɛ̃].

Possible reasons for the discrepancy in the results is that there is a risk of score inflations when using a quality rating task. The raters rarely used the lower number of the seven-point Likert scale, and although score inflation was consistent throughout the rating task and therefore not problematic for the analysis, it became more of an issue when comparing the results of the quality rating task to those of the forced-choice identification tasks which did not leave any option for inflation. On the other hand, the forced-choice identification task was not exempt from drawbacks. First, the total accuracy scores in this study might be lower than if only two sounds had been contrasted, which is often the case in the previously cited studies, as it was somewhat cognitively more demanding as the raters have to focus on three sounds. In addition, the fact that there was no possibility to select a "none of these sounds" option forced the raters to select one option by default even if the sound produced by a L2 participant did not correspond to any of the three vowels.

In conclusion, this study contributes to the growing literature showing that AV perceptual training not only leads to improvement in speech production, but also leads to greater improvement than auditory-only training (Hardison, 2003; Hazan et al., 2005). This has strong practical implications for computer-assisted language learning and should be further investigated. In addition, this study raised methodological concerns about the assessment of speech production by showing that different results can be obtained from different types of rating.

ACKNOWLEDGMENTS

The author would like to thank Drs. Debra Hardison, Aline Godfroid, Shawn Loewen, and Anne Violin-Wigent for their advice during this research project. This research was supported by a *Language Learning Dissertation Grant*, and both a *Dissertation Completion Fellowship* from the College of Arts and Letters and a *Research Enhancement Award* from the Graduate School at Michigan State University.

ABOUT THE AUTHOR

Solène Inceoglu is Assistant Professor in the Department of Modern Languages and Cultures at Rochester Institute of Technology (Rochester, NY) where she teaches French, French Linguistics, and German Culture. She received her Ph.D. in Second Language Studies from Michigan State University in 2014 where she taught courses on SLA, Language Teaching Method, and French Phonetics. Her research focuses on SLA, L2 speech perception/production, audiovisual L2 speech processing, co-speech gesture, and psycholinguistics. She has presented at AAAL, SLRF, Eurosla, New Sounds, ASA, AILA and CALICO. Contact information: scigsl@rit.edu

REFERENCES

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*(2), 339–355.

Benoît, C., Mohamadi, T., & Kandel, S. (1994). Effects of phonetic context on audio-visual intelligibility of French. *Journal of Speech and Hearing Research*, *37*(5), 1195–203.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics*, *61*(5), 977–985.

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

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, *20*, 1–16.

Erdener, D., & Burnham, D. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, *55*(2), 191–228.

Flege, J. E. (1988). The production and perception of foreign language speech sounds. *Human Communication and Its Disorders: A Review*, *2*, 224–401.

Flege, J. E. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. In R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 157–232). Amsterdam: John Benjamins.

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, *106*(5), 2973–2987.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125–3134.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522.

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, *119*(3), 1740–1751.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*(3), 360–378.

Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*, 298–310.

Kluge, D. C., Reis, M. S., Nobre-Oliveira, D., & Bettoni-Techio, M. (2009). The use of visual cues in the perception of English syllable-final nasals by Brazilian EFL learners. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent research in second language phonetics/phonology: Perception and production.* (pp. 141–153). Cambridge Scholars Publishing.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. a., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227–247.

Lopez-Soto, T., & Kewley-Port, D. (2009). Relation of perception training to production of codas in English as a second language. *Proceedings of Meetings on Acoustics*, *6*(1), 1–15.

Major, R. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Erlbaum.

Major, R., & Kim, E. (1996). The similarity deferential rate hypothesis. *Language Learning*, *46*, 465–496.

Pisoni, D. B., & Remez, R. E. (2005). *The handbook of speech perception*. (D. B. Pisoni & R. E. Remez, Eds.). Oxford, UK: Blackwell Publishing.

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). London: Erlbaum.

Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 51–79). Malden, MA: Blackwell.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*(2), 212–215.

Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, *36*, 314–331.

Wang, Y., Behne, D. M., & Jiang, H. (2008). Linguistic experience and audio-visual perception of non-native fricatives. *The Journal of the Acoustical Society of America*, *124*(3), 1716–1726.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033–1043.