EFFECT OF MULTIMODAL TRAINING ON THE PERCEPTION AND PRODUCTION OF FRENCH NASAL VOWELS BY AMERICAN ENGLISH LEARNERS OF FRENCH

By

Solène Inceoglu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies – Doctor of Philosophy

2014

ABSTRACT

EFFECT OF MULTIMODAL TRAINING ON THE PERCEPTION AND PRODUCTION OF FRENCH NASAL VOWELS BY AMERICAN ENGLISH LEARNERS OF FRENCH

By

Solène Inceoglu

Face-to-face interaction often involves the simultaneous perception of the speaker's voice and facial cues (e.g., lip movements) making speech perception a multimodal experience (Rosenblum, 2005). Research in second language (L2) speech perception suggests that participants benefit from visual information (Hazan, Sennema, Iba, & Faulkner, 2005; Wang, Behne, & Jiang, 2008) and that perception training can transfer to improvement in production (Iverson, Pinet, & Evans, 2011) and can be generalizable to novel stimuli (Hardison, 2003). Most studies so far have investigated consonants and, despite a couple of studies looking at the multimodal perception of vowels (Hirata & Kelly, 2010; Soto-Faraco et al., 2007), little is known about the effect of multimodal training on the acquisition of vowels. More specifically, no study has looked at the contribution of visual cues in the perception and production of L2 French.

The aim of this study is to provide a better understanding of the role of facial cues by examining the effect of training on the perception and production of French nasal vowels by American learners of French. The following research questions guide this study: 1) Does Audio-Visual (AV) perceptual training lead to greater improvement in perception of nasal vowels than Audio-Only (A-only) training does? 2) Does AV perceptual training lead to greater improvement in production of nasal vowels than A-only training does? 3) Does perception accuracy vary in relation to consonantal context? 4) Is training generalizable to novel stimuli?

Sixty intermediate American learners of French were randomly assigned to one of the following training groups: AV, A, and control. All participants completed a production pretest

and posttest, and a perception pretest, posttest and generalization test. The perception tests (monosyllabic words with various consonantal contexts) were presented within three modalities and with two counterbalanced orders: AV, A, V or A, AV, V. During the three weeks between the pretest and posttests, the AV and A groups received six sessions of perception training.

The results of the perception task showed that, contrary to the control group, both the A-only and AV groups improved significantly from the pretest to the posttest, but that the differences between the AV and A-only groups were not statistically significant. When comparing each vowel in each of the three modalities, there was however a trend in favor of the AV training group. The analysis of the consonantal context revealed that, for both training groups, accurate perception of the vowel was higher when the initial consonant was a velar (occlusive non labial) and was lower with palatals (fricative labial). Training was also shown to be generalizable to new stimuli with novel consonantal contexts. In addition, although both groups improved at the production posttest, the oral production of the AV training group improved significantly more than the production of the A-only training group did, suggesting that AV perceptual training (i.e., seeing facial gestures) leads to greater improvement in pronunciation.

To learners of French all over the world,
and especially to Johannes

ACKNOWLEDGMENTS

Although I had always been interested in how people learn languages, I was only formally introduced to the field of second language acquisition during an exchange program at the University of Western Ontario when I took a class taught by Dr. Jeff Tenant. Although he doesn't know it, it is in his class that I decided that I would one day pursue a PhD in second language studies.

Now that my graduate studies are coming to an end, I would like to recognize my professors who have provided me with great learning and teaching opportunities, and continuous support. I am deeply indebted to my chair Dr. Debra Hardison for her support and encouragements over the last five years, for her countless letters of recommendations, and for her expertise in audiovisual speech perception. The idea of this dissertation originated in her L2 speech perception class and I'm lucky to have had her as a mentor. I would also like to thank the other members of my committee, Dr. Aline Godfroid, Dr. Shawn Loewen, and Dr. Anne Violin-Wigent for their feedback and support during this project.

This dissertation would not have been possible without the help of several people: Elodie Hablot who kindly let me video-record her (twice!), the French instructors who facilitated recruitment in their classes, and, above all, the students who participated in my study. Their enthusiasm, their willingness to come to the lab eight times (and not dropping out!), and their genuine interest in my study made them the best participants ever!

I also need to thank Hisham Abboud (from Cedrus) for his help in programming SuperLab, Russ Werner for his technical assistance, and acknowledge the financial support from a Language Learning Dissertation Grant, a Dissertation Completion Fund from the College of

Arts and Letters at Michigan State University, and a Research Enhancement Award from the Graduate School at Michigan State University.

I am grateful for the friendships that have developed throughout my years in the Second Language Studies program. I deeply enjoyed sharing this journey with my cohort: Hyojung Lim, Wen-Hsin (Kelly) Chen, Jimin Kahng, and Yeon Heo. Thanks for the mutual support, for the many lunches and dinners, and for teaching me about your cultures! Special thanks go to Soo Hyon Kim, Luke Plonsky, Ryan Miller, and Tomoko Okumo for their help regarding job searching and graduation. Thanks to all my other SLS/applied linguistics friends at Michigan State and elsewhere. I can't wait for the next conference where we can meet again and catch up! I would also like to thank my non-SLS friends at MSU who helped me balance academic and social life and with whom life in East Lansing, MI was so much better: Swasti Mishra, Chili Cai, Mahdi Moazzami, Sarah Mécheneau, Shinya Yuge, Aditya Singh, Irene Carbonell, Samaneh Rahimi, and Valentina Denzel. Thanks also go to my family members in France, Turkey, Germany, and the USA who supported me throughout my graduate studies and were always proud of me.

Finally, all of my most heartfelt thanks go to Johannes who encouraged me to pursue a PhD in the US even if it meant being 6,817km (4,236 miles) apart and supported me throughout the ups and downs of (graduate) life. *Danke, dass du immer für mich da bist.*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Adults often have difficulty perceiving and producing the sounds of a second language (L2), especially when these sounds are not present in their native language inventory or when they resemble too closely sounds in their first language (L1) (Best, 1995; Flege, 1995). Studies have reported that adult L2 learners perceive L2 sounds differently than monolingual native speakers of the target L2 do (Best & Tyler, 2007), and that the production of L2 learners diverges from the phonetic norm of the L2 (Flege, 1997). Whether adult speakers can eventually produce L2 sounds with native-like accuracy is still debated (see Birdsong, 2006). Nevertheless, studies have shown that auditory training can contribute to improvement in L2 speech perception, indicating that perceptual patterns are modifiable to a certain extent (e.g., Flege & MacKay, 2004; Iverson & Evans, 2009; Iverson et al., 2011; Logan, Lively, & Pisoni, 1991) and that improvement in perception sometimes transfers to improvement in production (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lopez-Soto & Kewley-Port, 2009).

In the past two decades, researchers have asserted that speech perception is a multimodal experience involving both auditory and visual information (Rosenblum, 2005) and have commented that L2 speech perception studies have largely disregarded the importance of input from the visual modality (Hardison, 2003; Kellerman, 1990). Findings from audio-visual speech perception studies point to the beneficial effect of visual information, such as lip movements, on speech comprehension, discrimination, and learning. Such influences have been reported in research on L1 speech development with infants, in L1 speech processing experiments with degraded speech and/or mismatched information (e.g., Alm, Behne, Wang, & Eg, 2009; Behne et al., 2007; Binnie, Montgomery, & Jackson, 1974; McGurk & MacDonald, 1976; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Sumby & Pollack, 1954), in studies on hearing impairment

(e.g., Bergeson, Houston, & Miyamato, 2010; L. Bernstein, Tucker, & Demorest, 2000; Erber, 1974; Lachs, Pisoni, & Kirk, 2001; Owens & Blazek, 1985), and in investigations of speech perception in L2s (e.g., Goto, 1971; Hardison, 1996, 2003, 2005a, 2005b, 2007; Hazan et al., 2006, 2005; Hirata & Kelly, 2010; Kluge, Reis, Nobre-Oliveira, & Bettoni-Techio, 2009; Lively, Logan, & Pisoni, 1993; Massaro & Light, 2003; Navarra & Soto-Faraco, 2007; Pereira, 2012, 2013; Soto-Faraco et al., 2007; Walden, Prosek, Montgomery, Scherr, & Jones, 1977; Werker, Frost, & McGurk, 1992).

Most of the aforementioned studies on audio-visual (AV) speech perception have investigated consonants, and little is known regarding AV perception of vowels by normal hearing adults, most particularly by non-native speakers. Considering the fact that vowels are generally less visually salient than consonants, this dissertation aims at exploring whether American learners of French would be able to use the visual cues available to them to distinguish French vowels. French nasal vowels are particularly suited for an investigation of the effect of AV cues on the perception and production of L2 vowels for several reasons. First and foremost, their differences are visually salient, as they are placed on a continuum from hyper-rounded to unrounded (Zerling, 1989). Second, more rounded vowels exist in French than in English, and the rounding is different in the two languages. Finally, nasal vowels are often problematic for L2 learners, and any positive effects from training could have relevant pedagogical implications.

This dissertation is organized in the following way: In chapter 1, I will review relevant areas of the literature on speech perception and production. This will include a discussion of speech perception models, a review of previous empirical research on auditory and AV speech perception and production, a look at the effect of consonantal context on speech perception, and a synthesis of research on French nasal vowels. Chapter 2 describes the methodological designs

and implementation of the current training study, and chapter 3 reports the results of the empirical questions. In chapter 4, I will discuss the findings in light of the research questions. Finally, in chapter 5, I will summarize the findings of the study, discuss pedagogical implications, address some limitations, and make recommendations for future research.

CHAPTER 1: REVIEW OF THE LITERATURE

1.1 Models of cross-language speech perception

Several models of cross-linguistic speech perception have been developed to explain the perception and production of non-native sounds by L2 learners. The two most prominent models, the Perceptual Assimilation Model (PAM) (Best, 1994, 1995) and the Speech Learning Model (SLM) (Flege, 1995), both make predictions about the degree of difficulty for acquiring L2 phonemic contrasts and claim that the difficulties that listeners encounter learning their L2s are determined by the perceived similarities between their L1 and L2 phonetic systems.

Flege's Speech Learning Model (1995) suggests that L1 phonetic categories develop during childhood and are likely to block the formation of new categories for non-native consonants and vowels. In the original version of his model, Flege posited that L2 sounds are either classified as new, identical, or similar in relation to the L1 phonological categories. He later adopted the term of "perceived phonetic differences" between L1 and L2 phones or between two L2 phones and focused more on the perceptions of listeners rather than on the acoustic differences between the sounds. According to this model, L2 sound categories that are phonetically dissimilar to the native sound system are predicted to be easier to perceive and acquire due to the lack of interference between the two sound categories, and listeners will be more likely to create new categories for new L2 phones. Alternatively, if an L2 sound is similar enough to an L1 sound, it will be assimilated into an already existing L1 category. In some cases, however, because of "equivalence classification", an L1 phonological system will filter out important differences and assimilate L2 phones into existing L1 categories, resulting in misperception and accented production. Another important point from Flege (1987) is that L2 production is a reflection of the phonetic categories from an individual's perceptual experiences

and that L2 learners use articulatory gestures established during their L1 acquisition. According to the SLM model, accurately perceiving phonetic differences between two L2 phones will eventually lead to accurate production of these two L2 sounds. Conversely, Flege (1995, p. 239) notes that when a new L2 sound category is not created because of equivalence classification, "a single phonetic category will be used to process perceptually like L1 and L2 sounds (diaphones). Eventually, the diaphones will resemble one another in production."

The Perceptual Assimilation Model developed by Best (1994, 1995) also argues that problems in L2 speech learning lie mostly in perception, but it does not directly predict difficulties in production. The model predicts various degrees of discrimination difficulty for adult L2 listeners based on several patterns of assimilation. In the Single Category assimilation pattern, discrimination is expected to be poor because contrastive L2 segments are assimilated as good instances of the same L1 segments. This has been shown to be the case for Japanese ESL learners assimilating both the English /r/ and /l/ to the Japanese /r/ (Yamada & Tohkura, 1992). On the other hand, in the Two Category assimilation pattern, discrimination is expected to be excellent because each L2 segment is assimilated to a different L1 category. For instance, Best and Strange (1992) showed that Japanese /w/ and /j/ were assimilated to their American English counterparts /w/ and /j/. Good to moderate discrimination is also predicted when contrastive L2 segments are assimilated into the same L1 category but differ in terms of the goodness of fit to the category, one being a better instance of the category than the other (Category Goodness). An example of this assimilation pattern shows that French front and back rounded vowels were both assimilated to American English back vowels, but the French back vowels were less well assimilated (Levy, 2009a). Best (1995) also proposed that L2 sounds can be assimilated as uncategorizable speech sounds—when the L2 segments lie within the L1 phonological space but

cannot be assimilated to a L1 phonemic category—or can even not be recognized as speech sounds and therefore not be assimilated to speech.

In summary, both SLM and PAM posit that the perception issues that L2 learners encounter are due to the assimilation of L2 segments into L1 categories. The PAM is designed to investigate the perceptions of inexperienced L2 listeners, whereas the SLM predicts perception and production for more experienced L2 learners.

1.2 Cross-language studies on auditory speech perception

1.2.1 Previous studies

Studies on auditory speech perception indicate that non-native speakers have difficulty perceiving some L2 contrasts (Flege, 1995; Goto, 1971; Werker & Tees, 1984) even when they have been immersed in L2 environments for long periods of time (Flege & MacKay, 2004; Munro, Flege, & Mackay, 1996). A reason for this appears to be that L2 listeners have less well-developed phonetic categories due to differences in the quantity and quality of L2 input. Results, however, should be interpreted with caution as large differences have been found in perceptual experiments with L2 sounds. Factors influencing these individual differences include L2 language proficiency and language experience (Best & Strange, 1992; Best & Tyler, 2007; Bohn & Flege, 1990; Levy, 2004), the extent to which an individual keeps using their L1 (Flege & MacKay, 2004; Flege, 2002), the motivations of L2 learners (e.g., Bongaerts, van Summeren, Planken, & Schils, 1997; Flege & MacKay, 2004; Flege, 1988; Skehan, 1991), and the methodology used during the experiment to measure perceptual ability (Flege, 2003; Mack, 1989).

Studies have established that L2 perception in adverse listening conditions (e.g., when perception is degraded because of noise) is poorer for non-native listeners than for native

listeners (Cutler, Garcia Lecumberri, & Cooke, 2008; Cutler, Smits, & Cooper, 2005; Florentine, Buus, Scharf, & Canevet, 1984; Garcia Lecumberri, Cooke, & Cutler, 2010; Garcia Lecumberri & Cooke, 2006; van Dommelen & Hazan, 2010, 2012) even when non-native speakers performed similarly to native speakers in quiet conditions (Nábělek & Donahue, 1984; Takata & Nábělek, 1990). One reason is that native speakers make better use of contextual cues (Golestani, Rosen, & Scott, 2009; Mayo, Florentine, & Buus, 1997). Garcia Lecumberri and Cooke (2006) examined the perceptions of English and Spanish listeners of English intervocalic consonants accompanied by three types of noise and found that the noise affected the L2 listeners significantly more than it did the L1 listeners, suggesting that "non-native phonetic category learning can be fragile" (p. 2445).

Counter evidence has, however, been found across various studies. For instance, Cutler, Weber, Smits, and Cooper (2004) presented monosyllabic CV and VC nonsense words spoken by an American English speaker to American English listeners and to Dutch listeners and used three levels of noise (e.g., little, mild, and moderate). Their results showed that both vowels and consonants were consistently identified less accurately by the L2 speakers than by the L1 speakers, but that the performance asymmetry between the two language groups remained roughly constant across the three levels of noise. The fact that the non-native disadvantage in performance was approximately the same for each noise level suggests that noise did not affect L2 listeners more than it did L1 listeners. These results were similar to a study by Bradlow and Bent (2002) that revealed that L2 listeners were not more adversely affected by increasing levels of noise than were L1 listeners. The authors, however, noted that the results might have been affected by the fact that L2 listeners exhibited a floor effect for the -8 dB signal-to-noise ratio (SNR) condition. In a subsequent study, Cutler et al. (2008) established that the task used in the

2004 study mentioned above was the reason for discrepancy. This time, larger noise effects on consonant identification were observed for Dutch learners of English than for native-speakers of English.

In addition, research has suggested that speech perception differs not only between native-speakers and non-native speakers but also between monolingual speakers and bilinguals. For example, Spanish/English bilingual speakers who learned English before the age of six and had no noticeable foreign accent were found to score lower than monolingual English speakers on a word recognition test in noisy and noisy-with-reverberation conditions but not in quiet conditions (Rogers, Lister, Febo, Besing, & Abrams, 2006). These results were similar to a previous study that investigated three groups of bilinguals and found that early bilinguals (i.e., who learned their L2s during infancy and toddlerhood) performed better than late bilinguals (i.e., who started learning their L2s after puberty) but less well than monolingual speakers in noisy conditions. In quiet conditions, however, early bilinguals and monolinguals performed similarly (Mayo et al., 1997).

Recently, van Dommelen and Hazan (2012) further explored the effects of noise by investigating L2 speech intelligibility with a much larger sample of talkers than previous studies had used (i.e., 45 talkers including adults and children). Their results confirmed previous findings that intelligibility rates were significantly lower for L2 listeners, and they showed that factors determining intrinsic talker intelligibility were relatively language-independent. Moreover, the authors not only investigated the effect of noise but also compared the perception of words presented individually versus the same words presented as triplets, with the aim of finding whether L2 speech perception imposes a greater cognitive load on L2 listeners. Their

results suggested that "increased memory demands in speech perception may imply an increased cognitive load causing reduction in performance" (p. 1698).

1.2.2 Training studies

Of considerable theoretical interest to second language acquisition (SLA) researchers and also of pedagogical significance is the extent to which adult L2 learners can modify their perceptual patterns through training. Results of training experiments have indicated that, despite difficulties in cross-linguistic speech perception and even after the so-called Critical Period has passed, adult L2 learners still have the necessary auditory ability to distinguish L2 speech sounds, suggesting that language-specific perceptual patterns are modifiable to some extent (Best & Strange, 1992; Bradlow et al., 1997; Logan et al., 1991).

Factors to take into account when designing a training experiment are the type of training tasks, the stimuli used during testing and training, the duration of training, and the number of talkers presented during testing and training. Training types can involve either discrimination tasks, in which the listener must determine whether the stimuli presented are the same or different, or identification tasks, where the stimuli are presented in a forced-choice paradigm, either using a fading technique (see, Jamieson & Morosan, 1986; Morosan & Jamieson, 1989) or a high-variability phonetic method, which involves minimal pairs contrasting the sound under investigation in various phonetic environments. Training involving various phonetic contexts has been shown to enhance long-term modification of L2 learners' phonetic perceptions (Iverson, Hazan, & Bannister, 2005; Logan et al., 1991; Pruitt, Jenkins, & Strange, 2006). On the other hand, an early study on the acquisition of /r/ and /l/ by Japanese ESL learners showed that training incorporating stimuli from only one context did not enable L2 learners to modify their

phonetic perception or to accurately identify stimuli presented in novel phonetic environments (Strange & Dittmann, 1984).

Although no fixed standard exists regarding the duration of training, previous studies have reported training lasting from one session (Logan & Pruitt, 1995; Wang & Munro, 1999) to long term training over 45 sessions (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow et al., 1997). Generally though, researchers tend to implement training ranging from five to eight sessions (e.g., Iverson & Evans, 2009; Iverson et al., 2011; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Lengeris & Hazan, 2010; Rochet, 1995; Wang, Spence, Jongman, & Sereno, 1999). Regarding stimuli, researchers have used both synthetic tokens, which allow the exaggeration or reduction of differences along a continuum (Lengeris & Hazan, 2010; Rochet, 1995; Strange & Dittmann, 1984), and natural stimuli, which are more representative of real speech (e.g., Bradlow et al., 1997; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Lengeris & Hazan, 2010; Logan et al., 1991; Pruitt, Jenkins, & Strange, 2006).

Another methodological difference concerns the number of talkers used during testing and training. Studies have reported using one talker (Lively et al., 1993), four talkers (Wang et al., 1999), and five talkers (e.g., Bradlow et al., 1997; Lambacher et al., 2005; Lengeris & Hazan, 2010; Pruitt et al., 2006; Shin & Iverson, 2013) during training, and up to 10 talkers during testing (Iverson et al., 2011). The purpose of using several talkers during training is to strengthen L2 learners' abstract representations of L2 phones to further enable generalization when exposed to new talkers and novel stimuli. For instance, Japanese adults receiving perceptual training on the English /r-l/ contrast produced by a single talker failed to generalize their learning to novel stimuli despite improving from pretest to posttest (Lively et al., 1993). In general, findings have

shown that training that uses multiple talkers promotes generalization when exposed to new talkers who were not present during training, and these findings seem to indicate that L2 learners were able to establish robust categories (Huensch, 2013; Pruitt et al., 2006; Wang et al., 1999). However, some studies have reported that trainees performed better with familiar talkers than with new talkers (e.g., Lively et al., 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan et al., 1991).

1.2.3 Effect of perceptual training on production

A general consensus exists that most individuals who learn L2s as adults speak them with foreign accents (e.g., Derwing & Munro, 1997; Flege, Munro, & MacKay, 1995; Major, 2001). Factors leading to difficulties in accurately producing non-native sounds include the influence of L1s (e.g., Flege, Yeni-Komshian, & Liu, 1999; Piske, MacKay, & Flege, 2001), L2 input quality and quantity (e.g., Cummins, 1981; Flege & Liu, 2001; Oyama, 1976; Stevens, 1999), maturational constraints and age of acquisition (e.g., Flege, 1992; Lenneberg, 1967; Long, 1990; Munro et al., 1996; Patkowski, 1990; Scovel, 1988), motoric difficulties (Flege, 1987; Sapon, 1953), psychological factors and motivation (e.g., Bongaerts, van Summeren, Planken, & Schils, 1997; Gardner, Masgoret, Tennant, & Mihic, 2004; Moyer, 1999; Purcell & Suter, 1980), word familiarity (Trofimovich, Baker, Flege, & Mack, 2003), and orthography (Flege, 1991b).

Understanding the relationship between speech perception and speech production has been a long-standing concern for speech theorists and L2 researchers. As mentioned earlier, the SLM (Flege, 1987) posits that accurate perception of L2 phones and establishment of novel phonetic categories will eventually lead to accurate L2 speech production. Proponents of the Motor Theory (Galantucci, Fowler, & Turvey, 2006; Liberman & Mattingly, 1985; Liberman & Whalen, 2000; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970) have proposed an even

more direct link between speech perception and production and have argued that a specialized phonetic module exists that represents speech in terms of articulatory gestures that mediates both speech perception and speech production.

Evidence has shown that the degrees of accuracy in perceiving and producing L2 phones are related (Flege, Bohn, & Jang, 1997; Flege, 1988; Levy, 2009a, 2009b). Rochet (1995) examined the relationship between perception and production of the French vowel [y] by L1 Brazilian Portuguese and Canadian English speakers. In both production and perception tasks, the first group systematically substituted [y] with [i] and the second group substituted [y] with [u], suggesting that accented productions by L2 speakers might be perceptually motivated. Flege (1999) revealed that discrimination scores (perception task) of Italian learners of English correlated with their intelligibility scores (production task). Levy compared the results of a discrimination study of French vowels by American learners of French (2009b) to the results of a perceptual assimilation study (2009a) and found that in both studies, participants confused or assimilated front rounded vowels primarily to back rounded vowels. Studies have also indicated that L2 sounds that are perceptually difficult to distinguish and acquire also cause difficulty in production (Bohn & Flege, 1992) and that accuracy in both perception and production varies as a function of L1 (Flege et al., 1997) and language experience (Flege, MacKay, & Meador, 1999).

An important theoretical and pedagogical question concerns the extent to which perceptual training can be transferred to improvement in production. Results of auditory training studies have shown that knowledge gained during perceptual learning of L2 sounds transferred to the production domain (Bradlow et al., 1999, 1997; Lambacher et al., 2005; Lopez-Soto & Kewley-Port, 2009). For instance, a perceptual high-variability training study of the English /r/ and /l/ by Japanese participants revealed that the knowledge gained about the L2 liquid contrast

during training resulted in an improvement in production, suggesting that auditory training led to the creation of more accurate, gesturally defined phonetic categories (Bradlow et al., 1997). In a follow-up study, improvement in production after perceptual training was also found to be retained even three months after the experiment (Bradlow et al., 1999). In addition, Lambacher et al. (2005) investigated the effects of a highly variable identification training procedure with immediate feedback on the perception and production of the American English mid and low vowels by Japanese speakers. Their results showed that the performance of the participants in the training group (but not those in the control group) improved in the perceptual identification task, and that this positive effect was transferred to production. Participants in the training group were found to be more intelligible than those in the control group at the posttest, and an acoustic analysis revealed that the vowel categories of the training group had less spectral overlap than those of the control group. Finally, Lopez-Soto and Kewley-Port (2009) explored the effects of a three-session perceptual training on the production of English codas by Spanish speakers who had been residing in the US for less than ten years. Results revealed that the productions of participants substantially improved for consonants that were not accurately produced during the pretest. In addition, the authors noted a relationship between large gains in perception and large improvements in production.

The relationship between perception and production is still not clear, and the conclusions offered so far have been sometimes contradictory. On the one hand, researchers have suggested that the development of speech perception precedes development of production, and that accurate perception is a prerequisite for accurate pronunciation (Escudero, 2005; Flege, 1991a; Kleber, Harrington, & Reubold, 2011; Rochet, 1995). Flege (1989) noted that although the speech perception of L2 learners can be native-like, "it may take [them] some time to learn how

to produce sounds according to the plan encoded in phonetic representations" (p. 264). For instance, a study with Korean learners of English showed that their phonemic identification of English /r/ and /l/ was more native-like than their production (Borden, Gerber, & Milsark, 1983). On the other hand, some evidence exists that speech production can also precede speech perception for L2 learners (Bohn & Flege, 1997; Gass, 1984; Sheldon & Strange, 1982; Yamada, Strange, Magnuson, Pruitt, & Clarke III, 1994) and for bilinguals (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; Mack, 1989). Some of the participants in Sheldon and Strange's (1982) study were able to accurately produce the English /r/ and /l/ but were unable to reliably identify the liquid contrast. The same pattern of results was found in a later study investigating the production and perception of the English /w/, /r/, and /l/ liquids by Japanese participants with various degrees of language experience (Yamada et al., 1994). The authors found that the production abilities of some of their participants were better than the perception abilities of those participants, but the opposite was never the case. Gass (1984) also observed that L2 learners' production of bilabial /p/ and /b/ was native-like, whereas their perception did differ from that of native speakers, indicating that speech perception and production followed nonparallel developments.

Differences in results across studies may be due to the use of different testing materials and rating procedures (Flege, MacKay, et al., 1999), different cognitive demands of the tasks (Strange & Shafer, 2008), or individual variations in perception and production (Bradlow et al., 1997; Fox, 1982). Furthermore, from both a theoretical and methodological aspect, it is important to note that cross-modal comparisons of linguistic behavior are difficult to make because, as Mack (1989, p. 198) noted, "speech perception requires methodologies, task demands, and measurement and evaluation procedures that are inherently different from those

used in tests of speech production". Differences across modalities might therefore also be due to differences in the methods used rather than or in addition to differences between perception and production abilities.

1.3 Multimodal speech perception

All the studies mentioned in the previous sections look at speech perception as a primarily auditory phenomenon. However, theorists propose that speech perception is a multimodal process, involving the integration of auditory information (hearing) and visual cues (lipreading) (Rosenblum, 2005). In face-to-face conversation, speech perception is influenced by the actual sound of speech, as well as the facial and lip movements of speakers (Sumby & Pollack, 1954). The importance of the visual modality is supported by studies showing that the motion of vocal tract articulators correlates with facial movements (Jiang, Alwan, Keating, Auer Jr, & Bernstein, 2002; Yehia, Rubin, & Vatikiotis-Bateson, 1998). In a study using motion markers, Jiang et al. (1998) estimated that "about 80% of the variance found in vocal-tract movements can be estimated from the face" (p. 23). Evidence supporting the integration of auditory and visual modalities comes from the results of behavioral and neurophysiological experiments.

Visual cues have been shown to be particularly useful for enhancing the speech perception of hearing impaired individuals (Owens & Blazek, 1985), listeners with cochlear implants (Bergeson et al., 2010; Bergeson, Pisoni, & Davis, 2003; Lachs et al., 2001; Schorr, Fox, van Wassenhove, & Knudsen, 2005; Strelnikov et al., 2013) or the profoundly deaf (Erber, 1971, 1974). Visual speech also facilitates comprehension for listeners with good hearing ability in environments degraded by background noise (Benoît, Mohamadi, & Kandel, 1994; MacLeod & Summerfield, 1990; Rosenblum, Johnson, & Saldaña, 1996; Sumby & Pollack, 1954;

Summerfield, 1979), when messages are conceptually difficult to understand (Arnold & Hill, 2001; Reisberg, McLean, & Goldfield, 1987), and when listening to accented speech (Reisberg et al., 1987). For instance, Reisberg, McLean, and Goldfield (1987) found that English L1 speakers learning French were better at shadowing French sentences presented in an AV mode than they were at those presented only auditorily. Visual cues have also been shown to enhance the comprehension of lyrics when listeners watched singers' lips (Jesse & Massaro, 2010) and to influence the language development of infants, as visually impaired children with normal hearing acquire visually salient but auditorily difficult phonetic distinctions (e.g., /m/ vs. /n/) later than sighted children (Mills, 1983, 1987). In addition, although congenitally blind people produce accurate sounds, studies have shown that sighted speakers produce vowels which are further apart on the vowel space (Ménard, Dupont, Baum, & Aubin, 2009). In summary, seeing information about articulation can give people an advantage in perception and production and at the right SNR, visual information can make the difference between comprehension and incomprehension for listeners. In addition, the fact that visual cues have been shown to be beneficial even when speech was clearly presented (Jesse & Massaro, 2010; McGrath & Summerfield, 1985) seems to provide evidence that multimodal perception is the primary mode of speech perception, whereas auditory-only speech perception is an impoverished version of the information (Rosenblum, 2005).

One of the most famous examples demonstrating the contribution of both oral and visual information to speech perception is probably the "McGurk effect" (McGurk & MacDonald, 1976), where, among other stimuli, an audio /ba/ was dubbed onto a visual /ga/ and was perceived by native speakers of English as a /da/. Interestingly, even when people are aware of the illusion and are asked to focus on only one information channel, the effect of the visual input

on the audio stimulus remains (Massaro, 1987). The effect has stimulated much research and replication, and it shows how discrepant visual information influences auditory speech perception of consonants in various L1s (for French, see Cathiard, Schwartz, & Abry, 2001; Colin, Radeau, & Deltenre, 1998; Colin, Radeau, Deltenre, & Morais, 2001; for English, see Burnham & Dodd, 2004; Green & Gerdman, 1995; Green, Kuhl, Meltzoff, & Stevens, 1991; MacDonald & McGurk, 1978; Massaro, Cohen, Gesi, & Heredia, 1993; Walker, Bruce, & O'Malley, 1995; for Japanese, see Sekiyama & Tohkura, 1991; for Finnish, see Sams et al., 1998; for Dutch, see Gelder, Bertelson, Vroomen, & Chen, 1995; for Italian, see Bovo, Ciorba, Prosser, & Martini, 2009), various L2s spoken by native-speakers with different L1s (Fuster-Duran, 1996; Grassegger, 1995; Hardison, 1996; Hayashi & Sekiyama, 1998; Massaro et al., 1993; Sekiyama & Tohkura, 1993; Sekiyama, 1997; Werker et al., 1992), and to a lesser extent vowels (Massaro & Cohen, 1993). For instance, Sekiyama and Tohkura (1991) investigated the extent to which the McGurk effect is observable in other languages and how it is affected by noise. Ten Japanese native speakers were asked to listen/watch ten Japanese consonants (C + vowel /a/) in noise-free and noisy conditions. Results showed that the McGurk effect was small in the noise-free condition but stronger when noise accompanied the audio stimuli, suggesting that Japanese speakers make less use of visual information than English speakers. The same results were found with native speakers of Chinese living in Japan who were tested in Japanese and English (Sekiyama, 1997). A possible explanation suggested by Sekiyama to account for the difference in the results between American English-speaking participants and her participants is that in the Chinese and Japanese cultures, people tend to avoid gazing at the faces of speakers and are therefore less sensitive to visual information.

This interpretation, however, has been challenged by Massaro, Cohen, Gesi and Heredia (1993) who explained that the variations were due to the inventory of linguistic prototypes. In their study, Massaro et al. (1993) observed that Spanish, Japanese, and American English participants were susceptible to the McGurk effect, thus arguing that the underlying mechanisms for speech perception are similar across languages. In addition, the McGurk effect has been found to occur in Finnish syllables, isolated words, and words within sentences (Sams et al., 1998), in French (Colin et al., 1998; Werker et al., 1992), in Dutch (Gelder et al., 1995), in Italian (Bovo et al., 2009), in a cross-linguistic German/Spanish study (Fuster-Duran, 1996), and in a cross-linguistic German/Hungarian study (Grassegger, 1995).

The argument for a multimodal speech primacy illustrated by the results of the aforementioned behavioral experiments is supported by data from neural imaging experiments. In a study using magnetoencephalographic recordings, Sams and his colleagues (1991) attempted to identify in which area visual and audio speech are integrated and suggested that "visual information from articulatory movements has an entry into the auditory cortex" (p. 141). In an experiment using functional magnetic resonance imaging (fMRI), Calvert et al. (1997) demonstrated that the mere fact of observing lip movements, without having access to any speech sounds, was shown to activate the auditory cortex, reinforcing the idea that "seen speech" influences "heard speech". A potential problem in their method is that fMRI generates noise during image acquisition that might be interpreted as speech. In a follow-up study (MacSweeney et al., 2000), the team replicated the study without acoustic noise and confirmed their initial claim that silent speechreading activates the auditory cortex in the same way that listening to speech does. Although some studies failed to find a link between visual speech perception and activation in the primary auditory cortex (Bernstein et al., 2002; Sekiyama, Kanno, Miura, &

Sugita, 2003), many researchers agree that the brain is organized around multimodal input (Callan, Callan, Kroos, & Vatikiotis-Bateson, 2001). A possible explanation for the divergence in results is individual differences in lipreading ability among participants. Ludman et al. (2000) found that the auditory cortex was activated during silent lipreading, but that the participants with the lowest lipreading ability showed significantly less activation in the superior gyrus (i.e., where the auditory cortex is located) and the middle temporal gyrus (the exact function of which is unknown).

Although the recent findings in the neurophysiology of speech mentioned above support the idea that multimodal speech is integrated at an early stage, a debate remains regarding the primitives of the speech perception function (Rosenblum, 2005). Some models claim that the objects of speech perception are acoustic events (Diehl & Kluender, 1989; Massaro, 1987; Stevens, 1989), whereas others claim that these objects are gestural in nature (Fowler, 1986; Liberman & Mattingly, 1985).

1.3.1 Theories of multimodal speech perception

The auditory enhancement theory proposed by Diehl and Kluender (1989) suggests that the identification of speech sounds is influenced by acoustic cues processed by the auditory system. According to this theory, contrasts between the sounds in a given phonological inventory are robust because phonological systems have evolved to maximize auditory distinctiveness. For example, voiced stops have shorter closure intervals than voiceless stops, and vowel systems are consistent with this principle of maximal perceptual distinctiveness (see Fowler, 1989 for a critique).

Similarly, the quantal theory developed by Stevens (1972, 1989) seeks to explain why certain sounds—and systems of sounds—are favored cross-linguistically by examining the

relationship between articulatory parameters (e.g., vocal-track configuration) and acoustic output. Stevens proposes that although acoustic output changes when articulatory parameters are modified, phonetic "regions" exist in which the relationship between an articulatory parameter and the acoustic output is not linear, and some small articulatory changes have much stronger effects than others. Stevens' main claim is that linguistic contrasts involve differences between what he calls "quantal regions". For instance, that [i-u-a] are the three vowels preferred cross-linguistically (i.e., present in all languages) is explained because they are quantal vowels and, therefore, contrastive. Conversely, according to this theory, phonological systems that depend on distinctions within a same quantal region will be rare. Major criticisms to the theory are that it does not take into consideration distinctions that rely on several articulatory configurations and that it excludes temporal properties from the descriptions of features (for a critique, see Studdert-Kennedy, 1989).

The fuzzy-logical model of perception (FLMP) developed by Dominic Massaro (1987, 1998) posits that speech perception is a pattern recognition process and the result of the integration of all available sources of information. This model suggests that speech recognition follows three stages: 1) in the feature evaluation stage, an acoustic signal is analyzed in terms of auditory and visual features, but no one source of information alters any other source of information; 2) in the feature integration stage, the features of a given acoustic signal are matched against the features of prototypes stored in memory to determine which prototype best integrates the features of the signal; 3) in the decision stage, a sound is classified on the basis of the relative goodness of match between its features to those of the prototypes. In terms of bimodal speech perception, it is important to note that the FLMP suggests that auditory and visual signals are analyzed independently of each other and that their integration is rather late in

the process. The FLMP, therefore, belongs to the auditory class of speech perception theories, contrary to the two models that I will turn to now.

A prominent theory associated with the specialized speech mechanism view is the Motor Theory (Liberman & Mattingly, 1985), which postulates that articulatory gestures, rather than sounds, represent the fundamental unit of mental representation in speech (Liberman & Whalen, 2000). In other words, the primitives of speech perception are not the sounds but rather the articulatory gestures that produce these sounds. According to this view, the activation of motor representations is what constitutes speech perception, so that by knowing what the gestures are, one can tell the sets of words that have been produced. Another assumption of the model is that a phonetic module dedicated to speech perception and production controls the (co-)articulation of the gestures and enables the acoustic signal—if it is interpreted as speech rather than noise— to be rapidly and automatically converted to phonetic gestures. In this theory, an important distinction is made between two perceptual systems: the phonetic module deals with intended speech sounds, whereas the auditory module processes ordinary sounds or noise. The idea of a duplex system is, however, greatly disputed by studies in which perception of speech stimuli has been found to be similar to that of non-speech stimuli sharing critical temporal properties (Pisoni, 1977; Stevens & Klatt, 1974).

The direct-realist theory of speech perception (Fowler, 1986) is another theory that claims that speech perception is gestural in nature, but it stands in opposition to the motor theory in various ways. First, contrary to Liberman, Mattingly and their colleagues, the proponents of the direct-realist theory deny the existence of a speech module specialized for speech perception that would be separate from an acoustic module (see Fowler & Rosenblum, 1990, for evidence against duplex perception in an experiment on the perception of slamming doors). The term

"direct" in the name of the theory is meant to imply that the information in the acoustic signal is rich enough that no need exists for indirect mediation. Second, the direct-realist theory suggests that speech sounds are co-produced (Fowler & Smith, 1986) rather than co-articulated. The implication is that the speaker does not intend to co-articulate the sounds (i.e., motor theory) but that the sounds happen to be combined due to the mechanisms of speech production. Accordingly, the direct-realist theory suggests that because the overlap of vowels and consonants does not result in assimilation of gestures or merging of the two sounds, the vowels and consonants remain separate and independent events.

Despite disagreements on whether auditory and visual signals are processed independently before final determination or are combined early on in the speech perception process (Rosenblum, 2005), a consensus exists that access to visual information benefits speech perception. The following three sections will take a closer look at the role that the visual modality plays in speech perception and production. First, I will review studies on L1 and L2 AV speech perception with consonants, before turning to the emerging body of research on vowel perception and production, both in L1s and L2s.

1.3.2 Audio-visual (L2) studies on consonants

The majority of research on visual and auditory input in speech processing has investigated consonant sounds. Researchers have mostly focused on the aforementioned McGurk effect in English and other languages, the effects of noise (Alm et al., 2009; Binnie et al., 1974; Jiang, Chen, & Alwan, 2006; Ross et al., 2007; Sommers, Spehar, & Tye-Murray, 2005), and the effect of age differences in AV speech perception (Behne et al., 2007; Cienkowski & Carney, 2002; Massaro, 1984; Musacchia, Arum, & Nicol, 2009; Sommers, Tye-Murray, & Spehar, 2005).

Since Sumby and Pollack (1954) noted the advantage of seeing the faces of speakers when auditory messages are accompanied by noise, numerous studies have investigated the effect of various SNRs to better understand the principles of multisensory integration. For instance, Binnie and colleagues (1974) observed that in an auditory condition, voicing and nasality were less affected by noise than place of articulation, but when visual input was added to the audio signal, and most especially in poor SNRs, errors in determining the place of articulation decreased. In addition, participants in this study were able to perfectly recognize places of articulation in the visual-only condition.

One observation is that multisensory gain, or bimodal enhancement, tends to be strongest when the unisensory signals are at their weakest, which is to say that the contributions of seeing speakers' lip movements are most pronounced when auditory input is weakest. This property, highly influential in studies of multisensory integration in humans and other mammals, is known as inverse effectiveness (Stein & Meredith, 1993), and has been challenged by some studies (Manjarrez, Mendez, Martinez, Flores, & Mirasso, 2007; Ross et al., 2007). For instance, Ross et al. (2007) manipulated the SNR of the speech channel to investigate how much visual information facilitated speech comprehension. Their results suggest that intermediate SNRs (−12 dB) exist for which AV multisensory integration is enhanced, and not when auditory input is weakest as predicted by the inverse effectiveness principle. Therefore, the authors note that "the speech recognition system appears to be maximally tuned for multisensory integration at SNR levels that contain these minimal levels of input—that is, there is a window of maximal multisensory integration at intermediate levels" (p. 1152)

Alm and colleagues (2009) further expanded the study of the role of noise in AV speech perception by using a McGurk paradigm. They compared whether white noise and babble noise

influenced the use of auditory and visual information differently. Fifteen Norwegian participants were presented with congruent and incongruent AV stimuli containing stop-vowel syllables that varied in terms of place of articulation (labial, alveolar, velar + /a/) and voicing. Besides using two noise types, the authors also manipulated the level of noise by presenting the stimuli at 0 and −12 dB SNRs. Results showed that the place of articulation was more affected by white noise, whereas voicing identification was more affected by babble noise. The authors concluded that "an integrated, acute, and highly adaptable system of AVSP [audiovisual speech perception] is demonstrated by the way changes in the speech signal and acoustical context affect the contribution of auditory and visual information to speech perception" (p. 386).

A body of research has focused on age differences in AV speech perception and has revealed that younger adults demonstrate superior lipreading ability and perceptual skills than older adults (Cienkowski & Carney, 2002; Dancer, Krain, Thompson, & Davis, 1994; Sommers, Tye-Murray, et al., 2005; Spehar, Tye-Murray, & Sommers, 2004), probably due to an age-related decline in spatial working memory and information processing speed (Feld & Sommers, 2009). Reliance on visual cues in AV speech perception has been found to increase from infancy to young adulthood (Massaro, 1984) and from young adulthood to older age (Behne et al., 2007).

Similarly to research in L1 speech perception, most studies investigating AV L2 speech perception have focused on consonants, and more particularly on the contrast between problematic L2 sounds that can be absent in certain L1s. A tremendous amount of research has been conducted on English sounds, such as the /r/ and /l/ contrast (Goto, 1971; Hardison, 1996, 2003; Hazan et al., 2006, 2005; Lively et al., 1993; Massaro & Light, 2003; Walden et al., 1977), the fricatives (Wang et al., 2008; Wang, Behne, & Jiang, 2009; Werker et al., 1992), the

labial/labiodental contrast (Hazan et al., 2006, 2005), and the syllable-final nasals (Kluge et al., 2009).

The first study investigating the effects of AV information on L2 speech perception was conducted by Werker, Frost, and McGurk (1992). The authors presented L1 French and L1 English speakers with multimodal stimuli that consisted of conflicting auditory and visual information in which the auditory /ba/ was paired with either the visual /ba/, / va/, /ða/, /da/, /ʒa/, or /ga/. Their results showed that because French does not have interdental fricatives, French beginner and intermediate learners of English could not use the visible information effectively to accurately identify the interdental fricative. Instead, they tended to substitute /ða/ to /da/ or /ta/, demonstrating that they assimilated the interdental place of articulation with that of the closest French phoneme.

Hardison (1996) conducted the first investigation of the McGurk effect with L2 learners of English. Participants (L1s: Japanese, Korean, Spanish, and Malay) were presented with CV syllables where V was /a/ and C one of the following: /p, f, w, r, t, k/; in four conditions: AV, A, AV+noise, and A+noise. Results revealed a significant increase in the identification of /r/ and /r-f/ with visual cues for Japanese and Korean participants, respectively, but not for the Malay and Spanish participants, who already had similar phones. The study showed that AV perception was influenced by linguistic experience and that "the contribution of a given cue to the percept depends not only on its information value, but also on its value relative to another cue" (p. 56).

A three-week training study involving Japanese and Korean learners of English by Hardison (1998, 2003) investigated the influence of a talker's face and voice, the vocalic context, and word position on the perception and production of the American English /r-l/ contrast. Participants were divided into three groups: AV training, A-only training, and no training

(control group), and they completed two-alternative forced-choice tasks with the minimal pairs /r-l/ embedded in monosyllabic words with various phonetic contexts: initial singleton (road), initial cluster (crime), final singleton (heal), and final cluster (tires). For the pre-, post- and generalization tests, the stimuli were presented once in each modality: AV, A, and V. To evaluate the effect of perceptual training on production, participants' recordings of 100 words taken from the pretest stimuli were made at the pretests and at the posttests, and were later assessed by native speakers of English. Results indicated that the AV training group improved more than the A-only training group did and that perception was enhanced by visual information in contexts where the L1 phonology predicted difficulty. In addition, results of a generalization test involving new stimuli and a new talker showed that the training was generalizable and that improvement in production was significant, suggesting that improvement in perception transfers to improvement in production.

The advantage of AV input over unimodal auditory input in identifying different L2 consonants has been supported in a word identification study using the gating paradigm (Hardison, 2005a). Japanese and Korean speakers of English participated in the study and were tested on familiar, bisyllabic words beginning with /p/, /f/, /r/, /l/, and /s, t, k/ combined with high, low, and rounded vowels. The 32 Japanese and 32 Korean participants were evenly divided between four groups: AV-gated stimuli for participants receiving minimal pair training involving /r/-/l/, /p/-/f/, and /θ/-/ð/, AV-gated stimuli without training, A-gated stimuli with training (same stimuli as AV training), and A-gated stimuli without training. The posttest data showed that words were identified significantly earlier by both L1 groups following training, and identification was facilitated by visual information, indicating that visual speech cues have a

priming role on L2 word identification and that sensitivity to visual information in non-native sounds can be enhanced through AV training.

In another gating study, Hardison (2005b) investigated the influence of visual cues, speech style (unscripted vs. scripted), word length (one vs. two syllables) and initial consonant visual category on spoken word identification by L1 English and L1 Japanese participants. Results were consistent with her previous perception training study (Hardison, 2003), confirming the enhancing effect of visual cues on speech perception and word identification, especially for words with problematic sounds like /r/, /l/, and /θ/. In addition, excised bisyllabic words were identified earlier than monosyllabic ones, but this did not reveal an effect of speech style.

Similarly to Hardison (1998, 2003), Hazan et al. (2005) conducted two experiments investigating the effect of AV perceptual training on the perception and production of consonants by Japanese learners of English. In experiment one, they used nonsense words (pretest and posttest) and real words (training) with the consonants /p/, /b/, and /v/ embedded within the following structure: CV, VCV, or VC, where V was either /i/, /u/, or /ɑ/. For the pretest and posttest, these tokens were presented twice within three modalities, and two orders were used for the presentation of the three conditions: (AV, A, V) or (A, AV, V). Participants received ten sessions of AV or A-only training over a period of four weeks, and results demonstrated that the AV training group improved more in the perception of the labial/labiodental contrast than the A-only training group did. In addition, the data showed that the A-only training group showed little improvement in their use of visual cues, whereas the AV training group improved both in the AV and A modalities, suggesting that members of this group became more sensitive to the auditory information. Interestingly, and contrary to the authors' expectations, the participants who received higher scores in the V modality at the pretest and were assigned to the AV training

group did not improve more than the other participants in their training group. Experiment two investigated the effect of AV training on the less visually salient /r/-/l/ contrast. Stimuli were embedded in initial (CV, cCV) or intervocalic (VCV, VcCV) nonsense words with /k/ and /f/ as additional consonants. The pre/posttest and training procedures were the same as for experiment one, but the 62 participants were divided into three training groups: A, AV, and AV synthetic (i.e., the speech signal was synchronized with the articulatory gestures of the talking head "Baldi"(Massaro, 1998)). Contrary to Hardison (1998, 2003), the AV training group in this study did not improve more than the A-only training group did. Instead, all the three training groups significantly improved, but the A-only training group improved more in the A modality, while the AV training group was better at using visual cues. In terms of production, however, greater improvement was demonstrated by the AV group, suggesting that exposure to visual cues without any training in production can lead to improvement in pronunciation. The authors concluded that AV training is more efficient than A-only training when the visual cues to the contrast are salient enough.

With the goal of identifying sounds that might be particularly appropriate to AV training, Ortega-Llebaria, Faulkner, and Hazan (2001) investigated auditory-visual and auditory confusions of 16 British English consonants by 36 L1 Spanish speakers with various linguistic proficiency levels and compared them to those of twelve L1 English speakers. The consonants were embedded into CV, VCV, or VC words, where V was one of /i, ɑ, u/. The sound strings were presented via Baldi. The results showed that the presentation of both audio and visual information enhanced the perception of consonants by both the native and non-native participants. This was, however, not the case for the contrasts /v/-/b/ and /ð/-/d/, which were still

problematic for Spanish L1 speakers, probably because they used the visual cues as allophonic features.

Hazan et al. (2006) conducted two experiments to test the effects of visual cues on the perception of L2 consonant contrasts. In the first experiment, Spanish and Japanese learners of English were presented with the consonants /p/, /b/, and /v/ embedded within nonsense CV, VCV, or VC words, where the vowel was either /i, ɑ/ or /u/. The stimuli were presented in two orders of presentation: (A, AV, V) or (AV, A, V), with two blocks of 81 items per condition. The authors reported that Spanish speakers showed greater sensitivity to visual cues than Japanese speakers when tested on their perception of a labial/labiodental consonant contrast in English because the labiodental [v] does exist in Spanish as an allophone of /f/. They also observed that some Japanese L1 speakers were better than others at learning to associate visual cues with appropriate phoneme labels, suggesting that individual differences in sensitivity to visual cues are an important factor to take into account. In a second experiment, the authors tested Japanese and Korean learners of English on their perception of the consonants /l/ and /r/ embedded in CV, cCV, VCV, and VcCV nonsense words, where the vowels were the same vowels as in experiment one and the initial consonant of the consonantal clusters was either /k/ or /f/. The results showed that the AV condition was not beneficial for either group, probably because of the lack of saliency of the consonantal contrast, confirming that visual cues do not enhance speech perception for native speakers and non-native speakers in the same way.

In the same vein, Wang and her colleagues (2009) investigated the influence of three L1s backgrounds (Korean, Mandarin Chinese, and English) on the AV perception of English CV syllables containing fricatives with three places of articulation (labiodentals, interdentals, and alveolars). Their study design was similar to that used by Hardison (1996, 2003), with the

presentation of the stimuli in A-only, V-only, and AV modalities, but they also added an incongruent AV modality to tease apart the influence of auditory and visual information in a single stimulus and determine the contribution of each component modality. They noted that the 20 Mandarin speakers outperformed the 15 Korean speakers in the perception of the visual cues for the English labiodental fricatives, since Mandarin Chinese does contain labiodental fricatives while Korean does not. The results of the interdental fricatives, nonnative for both L2 groups, were better in the AV condition than in the A-only condition, confirming previous observations that listeners can use visual information in their L2 as an additional channel of input. In a similar study, Wang, Behne and Jiang (2008) further explored the effect of visual cues on the perception of English fricatives by looking at the role of linguistic experience and length of residence (LOR). They recruited 20 Mandarin Chinese speakers who had been living in Canada for an average of two years (short LOR) and 15 Mandarin Chinese speakers who had been in the country for about ten years (long LOR). Their results highlighted the positive effects of LOR and linguistic experience, as participants in the long LOR group approximated the pattern of the native speaker group in their use of AV information. The short LOR participants, on the other hand, relied on visual information as an additional source of input but could not interpret the visual cues accurately. The authors therefore suggested that "auditory learning may precede and is accompanied by visual learning, resulting in an effective integration of AV speech information" (p. 1724).

Following the methodology of Hazan et al. (2006), one small-scale study looked at the perception of English syllable-final nasals by Brazilian Portuguese learners (Kluege et al., 2009). The participants were presented with two blocks of 18 monosyllabic CVC words. The target words were minimal pairs contrasting /m/ and /n/ in the syllable-final position preceded by one

of the following vowels: /ɪ-ɛ-æ/. Similarly to previous research (e.g, Hardison, 2003; Hazan et al., 2006; Wang et al., 2009), the results of the perception test were better in the AV modality and thus confirmed the importance of visual input for the perception of a visually distinctive contrast. However, contrary to some studies (e.g., Hazan et al., 1996), the percentages of accurate identification were lower in the A modality than in the V modality, suggesting that the participants did rely on visual cues to contrast the two consonants. Analyses of the vocalic context showed that a low preceding vowel favored the identification of the alveolar coda nasal and a high preceding vowel disfavored the accurate identification of the English coda nasal bilabial.

Another methodological design enabling researchers to investigate the effect of AV speech perception with L2 learners involves conducting experiments with degraded speech. Perception of speech in adverse conditions has been found to be difficult, but it is particularly challenging for L2 speakers (e.g., Cutler, Garcia Lecumberri, & Cooke, 2008; Rogers, Lister, Febo, Besing, & Abrams, 2006; Takata & Nabelek, 1990). Hazan, Kim, and Chen (2010) conducted an experiment in which iterations of /ba/, /da/, and /ga/ by five Australian English and five Mandarin Chinese speakers were presented to Australian English, British English, and Mandarin Chinese participants. The stimuli were presented in four conditions: A-only, V-only, congruent AV, and incongruent AV. For the latter, three types of stimuli were prepared: (1) an auditory /ba/ dubbed onto a visual /ga/, (2) an auditory /da/ dubbed onto a visual /ba/, and (3) an auditory /ga/ dubbed onto a visual /ba/. To investigate the effect of visual and aural degradation, stimuli were presented either with or without blurring and, with or without noise, or with both noise and blurring. Their results showed that in the AV condition, when one channel of information was degraded, the other channel increased in influence, but that English L1

participants showed a stronger weighting of visual information for stimuli produced by non-native speakers than the L1 Chinese participants did. In addition, one important finding is that individual variation, independent of language background, played a great role in how stimuli were perceived and how visual information was weighted.

1.3.3 Audio-visual (L2) studies on vowels

The previous section reviewed work investigating the effect of AV cues on (L2) consonants. Although much work has been done with consonants, visual intelligibility of vowels has also received some attention despite that vowels might appear less visually salient than consonants. It is evident that only gestures that are produced with externally visible movements can be visually beneficial for perceivers to discriminate sounds (Summerfield & McGrath, 1984). The reason vowels are considered less salient than consonants is that the major determinant of the acoustical structure of vowels is the position of the body of the tongue, which is not always easy to observe (Stevens & House, 1955). Nevertheless, as Summerfield (1991) noted, "under optimal conditions, all English vowels are visibly distinct" (p. 119) due to a strong correlation between the height of the tongue in the mouth and the vertical separation of the lips.

Montgomery and Jackson (1983) videotaped four talkers producing 15 English vowels and diphthongs, and they measured the height, width, and area of lip opening during vowel production. They found that lip opening measurements were only moderately good predictors of perceptual vowel confusion and noted that significant differences across the four talkers prevented categorization of vowels simply based on lip openings. The authors and others (McGrath, Summerfield, & Brooke, 1984) noted that beside lip openings, other factors influencing the visual perception of vowels include the visibility of the teeth, the visibility of the tongue, temporal information on vowel production, the extent of lip rounding, and the degree of

lip protrusion. For instance, in a lipreading study investigating the role of the visibility of the teeth, McGrath et al. (1984) found that normally hearing subjects showed reduced use of the lip rounding dimension when the teeth were not visible and had therefore more difficulty perceiving the contrast /i-ɪ/. The visibility of the teeth has been shown to provide useful information regarding the place of articulation and helps distinguish contrasting vowels with similar lip shapes (Summerfield, MacLeod, McGrath, & Brooke, 1989).

Keeping in mind that vowel contrasts can be salient, research on AV intelligibility of vowels has looked at various languages, taking into consideration perceivers from different L1s and various types of vowel distinctions (Benoît et al., 1994; Johnson, Strand, & D'Imperio, 1999; Lisker & Rossi, 1992; Öhrström & Traunmüller, 2004; Ortega-Llebaria et al., 2001; Robert-Ribes, Schwartz, Lallouache, & Escudier, 1998; Summerfield et al., 1989; Summerfield & McGrath, 1984; Traunmüller & Öhrström, 2007; Valkenier, Duyne, Andringa, & Başkent, 2012).

For instance, Summerfield and McGrath (1984) tested whether the McGurk effect would occur in the AV perception of vowels. They carried out experiments in which auditory synthetic vowel continua (/i/-/u/, /i/-/a/, /u/-/a) in a [bVd] context, for instance /bad/, were paired with visual /u/, /a/ or /i/, for instance /bud/. Their results indicated that visual information biased the native English-speaking participants' perceptions of the vowels in the same manner as had been demonstrated with consonants. Vowels presented auditorily were identified as more like the visual vowels with which they were paired, especially in the /i/-/u/ continuum, even when participants were aware of the mismatch and were asked to respond to the audio stimulus only.

Integration of incongruent visual information was also observed in a study where participants were trained speech researchers. Lisker and Rossi (1992) asked their participants to

identify the degree of lip-roundness of ten rounded and eight unrounded French and non-French vowels presented aurally, visually, audiovisually, and in an incongruent AV condition. Overall, the judgments of the participants in the incongruent AV condition were affected by the visual information despite that they were asked to focus solely on the auditory signal. The probability of judging an unrounded vowel presented auditorily as unrounded increased when the visual information showed a rounded vowel. For example, the combinations [i_{audio}/y_{visual}] and [$e_{audio}/ø_{visual}$] elicited responses that were about 60% in favor of roundedness, while the reverse combinations [y_{audio}/i_{visual}] and [$ø_{audio}/e_{visual}$] resulted in responses that were less than 25% in favor of roundedness. The results, however, also revealed that individual differences in lipreading and types of vowels greatly influenced AV speech perception. Participants focused mostly on the auditorily signal, but some relied heavily on lipreading.

Green and Gerdeman (1995) further investigated the McGurk effect in an experiment with mismatched vowels. The authors dubbed auditory /ba/ and /bi/ tokens onto visual /gi/ and /ga/ stimuli, respectively, so that the AV stimuli not only conflicted in the initial consonant but also in the vowel. The results showed that when the visual and aural vowels matched, the McGurk effect was strong (about 75%), but in the vowel mismatch condition, the size of the McGurk effect was reduced to 44%, thus demonstrating that the participants were sensitive to the coarticulatory information between the consonant and its following vowel.

Johnson and colleagues (1999) conducted several experiments to investigate how the gender expectations of listeners affect AV speech perception of vowels. In particular, they looked at gender mismatch, gender voice stereotypicality, and gender abstractness (e.g., in one experiment, participants were asked to imagine faces). Participants were presented with tokens along a *hood-hud* continuum where the female and male acoustic stimuli were crossed with male

and female visual stimuli. Their results suggest that listeners' impressions of talker gender affected the location of phoneme boundaries.

In a study investigating the effects of phonetic context on three vowels in French, Benoît and his colleagues (1994) found that [a] was most auditorily intelligible, followed by [i] and then by [y], whereas [y] was most visually intelligible, followed by [a] and [i]. These results demonstrated AV complementarity in the perception of French vowels at the information level, but the size of their set of tested vowels was small. The results of another study testing seven French oral vowels [i, e, y, ø, u, o, a] suggest that, in the audio condition, the most robust identification cue was vowel height, followed by backness, and finally rounding, whereas in the video condition, rounding was more significant than height, and backness was almost invisible (Robert-Ribes et al., 1998).

Öhström and Traunmüller (2004, 2007) have conducted several studies on the AV perception of Swedish vowels. They presented nonsense /gVg/ syllables, where V was one of the following: /i, y, e, ø/, in auditory, visual, congruent AV, and incongruent AV conditions. Stimuli were recorded by two men and two women. Results from their 21 native participants showed that the McGurk effect occurred with Swedish vowels, as a visual [y] combined with an auditory [e] was mostly perceived as an [ø], and an auditory [y] combined with a visual [e] was mostly perceived as an [i]. The authors observed that the linguistic features were weighted differently according to the information medium, and that perception of a feature was based on the modality that provided the most reliable information. Thus, openness was mostly perceived auditorily, while roundedness was perceived visually. The results also supported previous evidence that men rely less on lipreading than women (Daly, Bench, & Chappell, 1996; Johnson, Hicks, Goldberg, & Myslobodsky, 1988).

A similar study was conducted with Dutch front vowels /i, y, e, ʏ/, which differ in terms of rounding and height (Valkenier et al., 2012). Besides using congruent AV stimuli and three types of incongruent AV stimuli (fully incongruent, incongruent height, and incongruent rounding), the authors also added background noise (30dB, 0dB, –6dB, –12 dB, and –18 dB) to enhance the reliance on visual information. Nonsense /χVχ/ syllables produced by a female speaker were presented to 16 native participants. Similarly to Öhström and Traunmüller (2004, 2007), rounding was mostly perceived visually, and height information was transmitted auditorily, but height, which is not a visually salient feature, was also found to be affected by visual information. Contrary to the authors' expectations, the results of the incongruent height condition were not similar to those of the congruent condition, indicating that participants were negatively influenced by the incongruent—yet less salient— information.

In comparison to the amount of studies on L2 consonants that have been conducted over the past two decades, relatively little work has been done regarding the AV perception of L2 vowels, and the few studies that have started to investigate AV vowel perception are very recent and limited in terms of generalizability and comparison (Hirata & Kelly, 2010; Navarra & Soto-Faraco, 2007; Pereira, 2012, 2013).

Navarra and Soto-Faraco (2007) recruited 50 Catalan/Spanish bilinguals raised in Catalan monolingual families and 53 raised in Spanish monolingual families. All participants were born and raised in Barcelona and were therefore exposed to the two languages from an early age, usually before three years old when they attended day-care or kindergarten. The authors found that, contrary to Catalan-dominant bilinguals, Spanish-dominant bilinguals could not distinguish the Catalan vowel contrast [ɛ - e] in an audio-only condition. However, when visual articulatory information was added, both groups could perceive the contrast, and when only visual

information (e.g., no sound) was presented, neither group could discriminate the two phonemes. These findings suggest that visual articulatory information enhances L2 speech perception at the level of phonological processing by way of multisensory integration. Although not only investigating the cross-linguistic perception of vowels, another study highlighted the importance of visual cues when discriminating languages and found that Spanish-Catalan bilinguals were able to distinguish the two languages in a visual-only condition, whereas Spanish monolinguals who knew only one of the languages were less accurate, and English and Italian participants who were unfamiliar with the two languages were unable to make the discrimination (Soto-Faraco et al., 2007).

Hirata and Kelly (2010) investigated whether multimodal input, namely information such as lip movements and hand gestures, helped improve the ability of native English speakers to perceive Japanese vowel length contrasts. Participants were divided into four training groups: (1) Audio-only, (2) Audio-mouth, (3) Audio-hands, and (4) Audio-mouth-hands, and each received four sessions of training between a pretest and a posttest that took place two weeks later. The stimuli were embedded in a carrier sentence and were produced by four speakers (two men and two women) who were different from the speakers in the pre/posttests. The results showed that all experimental groups improved from pretest to posttest, but the improvement was greatest when mouth movements accompanied the auditory training (i.e., Audio-mouth group). Seeing hand gestures did not facilitate L2 speech perception, in contrast to previous studies that demonstrated that hand gestures can facilitate the learning of new L2 vocabulary (Kelly, McDevitt, & Esch, 2009).

To my knowledge, the only other AV training studies investigating the perception of vowels have been conducted on eleven English monophthongs (Pereira, 2012, 2013). In the first

study, Pereira (2012) compared the effects of three types of training: A, AV, and V on the perception of English vowels by 47 native speakers of Spanish. Each group received five sessions of training before completing a posttest that tested them on the perception of the vowels in isolated words and on their sentence processing (i.e, true/false type of sentences). The results showed that all the groups improved from pretest to posttest but that they did not differ in the rate of improvement. In addition, the improvement in the perception of words in isolation did not transfer to the processing of words in sentences. Most importantly, no significant difference in improvement was found between the AV and A groups, suggesting that participants were not able to take advantage of the visual contrasts among English vowels. In a further study, Pereira (2013) tested 37 Spanish advanced learners of English and 20 native English speakers on their perception of English vowels in real CVC words. This time, each participant was tested in three conditions: A, AV, and V, with two orders of presentation, (A-AV-V) or (AV-A-V), counterbalanced across participants. In addition, the stimuli presented to the native speakers were accompanied by noise (-10 dB SNR) to prevent a ceiling effect. Results indicated that native English speakers performed better in the AV condition than they did in the A condition, demonstrating that they were able to rely on visual cues. On the other hand, quite similarly to her previous research, the researcher noted that the performance of the Spanish participants did not differ between the A and AV conditions. Although they could use the visual information to some extent when they were forced to in the V condition, they failed to integrate the visual and audio information in the AV condition.

1.4 Effect of consonantal context on speech perception

Some AV and lipreading studies have noted that adjacent vowels influence the perception of consonants (Benguerel & Pichora-Fuller, 1982; Benoît et al., 1994; Hardison, 2003; Owens &

Blazek, 1985; Sheldon & Strange, 1982; Son, Huiskamp, Bosman, & Smoorenburg, 1994). For instance, in an investigation of viseme[1] classification by hearing-impaired and normal-hearing viewers, Owens and Blazek (1985) remarked that consonants produced in VCV syllables with /u/ were less accurately perceived (average correct score of 21.5% between the two groups) than consonants produced with /ɑ/ (43%), /ʌ/ (39.5%), and /i/ (32.5%). In addition, the intelligibility of French consonants in auditory and the AV modalities was observed to be highest in the [a] context, followed by the [i] context, and finally the [y] context (Benoît et al., 1994). English consonants were more easily recognized visually in the [æ] and [i] contexts than they were when accompanied by the rounded vowel [u] (Benguerel & Pichora-Fuller, 1982). The Dutch vowel /a/ led to higher accuracy of performance in lipreading identification of consonants (Son et al., 1994), and the AV perception of the English /r/-/l/ contrast by Japanese and Korean learners was shown to be influenced by the vowel context (Hardison, 2003).

Surprisingly, only one study has investigated the effect of consonantal context on the speechreading accuracy of vowels (Montgomery, Walden, & Prosek, 1987). Thirty viewers were presented with symmetric CVC syllables where C was one of the following: /p,b,f,v,t,d,f,g/, and asymmetric CVC syllables in the following contexts: /hVg/,/wVg/, and /rVg/. The vowels used were /i-ɪ-ɑ-u-ʊ/, and stimuli were recorded by two female speakers. Results demonstrated that the most accurately identified vowel was /a/, the least accurately identified vowel was /u/ and, tense vowels were more easily identified than lax vowels. This was hypothesized by the authors to be due to vowel duration (i.e., tense vowels have longer visual cues) and the fact that consonantal coarticulatory influence is stronger on lax vowels. More importantly, consonant

[1] The term "viseme" is derived from "visual phoneme" and refers to any speech segment that is visually contrastive from another (Fisher, 1968). For instance, the bilabials /p, b, m/ are one viseme and they contrast with the viseme /v, f/.

effects varied across the two talkers. For talker 1, vowels were better identified in the following contexts: /pVp/, /bVb/, and /hVg/; and less accurately identified in /ʃVʃ/, /fVf/, and /wVg/. By contrast, /rVg/, /gVg/, and /wVg/ yielded better vowel identification accuracy for talker 2, and /pVp/, /ʃVʃ/, and /fVf/ were identified less accurately. To draw a better picture of the effect of consonantal context on vowel perception, the authors conducted a further analysis according to consonant features. Overall, the neutral /h-g/, Low Lab (i.e., a small visual labial component) /t-d-g/, and Stops (i.e., rapid release and quicker transitions into vowels) were found to be most helpful for identifying what was spoken by both talkers, and High Lab (i.e., more visible labial components), including fricatives and labiodentals, were identified as the least helpful consonantal contexts. The results are, however, to be taken with caution as vowel identification accuracy was greater for talker 2 in the /p/ and /b/ consonantal contexts, indicating important interactions between talkers and phonetic contexts.

Because so few AV studies on vowel perception exist, and since none of them investigated the consonant effect, whether adjacent consonants have an effect on the perception of vowels in AV studies remains an open question—one that is addressed in the current study. The literature on lipreading and auditory speech perception research is, however, a first step toward understanding the effect of consonantal context on vowel perception.

Previous research has demonstrated that the spectral characteristics of vowels (i.e., formant frequencies and trajectories) differ as a function of consonantal context (Lindblom, 1963; Stevens & House, 1963). Studies have shown that American English vowels in various consonantal contexts were more easily perceived than in isolation (Gottfried & Strange, 1980; Rakerd, Verbrugge, & Shankweiler, 1984; Strange, Edman, & Jenkins, 1979; Strange, Verbrugge, Shankweiler, & Edman, 1976; Yakel, 2000) indicating that “dynamic acoustic

information distributed over the temporal course of the syllable is utilized regularly by the listener to identify vowels" (Strange et al., 1976, p. 213). Final consonants were found to aid identification more than initial consonants (Strange et al., 1979), and closed vowels appeared to be more hindered by context than open vowels (Rakerd et al., 1984). On the other hand, some studies failed to find differences in vowel identification between isolated vowels and vowels in consonantal context (Assmann, Nearey, & Hogan, 1982; Diehl, McCusker, & Chapman, 1981; Macchi, 1980). Diehl et al. (1981) suggested that several non-perceptual factors, such as memory load and how participants are asked to provide their responses, might be accountable for the differing effects of consonantal context. Finally, other studies suggest that consonantal context affected vowel perception (Gottfried, 1984; Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001). L2 learners of French had more difficulty discriminating certain vowels when they were in the consonantal context /tVt/ as opposed to when they were presented in isolation (Gottfried, 1984), for which the author attributed the cross-linguistic variances to differences in phonotactic constraints and phonological differences. Also, of particular interest for the current study, Gottfried (1984) noted that front rounded vowels were equally difficult to identify and discriminate in all contexts.

More recent studies (Levy & Strange, 2008; Levy, 2009a, 2009b) further investigated the discrimination and assimilation of Parisian French vowels by American English learners of French with various linguistic experiences. Levy and Strange (2008) conducted a study on the discrimination of the vowels /y, œ, u, i/ embedded in the sentence "j'ai dit neuf /raCVC/ à des amis" (i.e., I said nine /raCVC/ to some friends) produced by three female native speakers, where the CVC was either /bVp/ or /dVt/. The stimuli were presented in a categorical AXB discrimination task to experienced learners of French and individuals with no experience with

the language. Results revealed that the consonantal context did not influence vowel perception for the experienced group. Participants were good at discriminating all vowel pairs except /y-u/, which they confused in both contexts. Conversely, inexperienced listeners confused all the pairs equally and were influenced by the consonants. The bilabial context led to lower scores in the discrimination of /i-y/ than the alveolar context did, whereas the reverse pattern was found for /u-y/, /y-œ/, and /u-œ/. In a follow-up study, Levy (2009b) extended her analysis by including a third group of participants with moderate experience with French (i.e., formal classroom instruction but no immersion experience) and additional vowel contrasts (i.e., front rounded vs. back rounded /y-o/ and /œ-o/; and front rounded vs. front unrounded /y-ɛ/, /œ-ɛ/, and /œ-i/). Results confirmed that the consonantal context influenced L2 speech perception. The previously investigated front vs. back rounded vowels and the additional /œ-o/ contrast were again found to be less accurately perceived in the alveolar context than in the bilabial context, as was the discrimination of front rounded vowels from each other. Levy concluded by emphasizing that "perceptual training protocols that take consonantal context into consideration might better assess listeners' perceptual difficulties with vowels and gain effectiveness by targeting those contexts in which listeners have the most difficulty" (p. 2681).

Consonantal context has also been shown to affect the assimilation of vowels by L2 listeners. The same participants as those described in Levy (2009b) were asked to classify French vowels in terms of six American English vowel categories and rate them for goodness of fit (Levy, 2009a). Results indicated that the assimilation patterns of participants were affected by consonantal contexts. For example, participants assimilated /y/ more often to the American English /ju/ in the bilabial context than they did in the alveolar context, supporting the PAM's predictions (Best, 1995) that "the consonants surrounding vowels affect to which native vowel a

non-native vowel will be assimilated and the goodness of fit to that category" (Levy, 2009b, p. 1150). Similarly, studies on Japanese learners of English also showed that consonant place of articulation (i.e., bilabial, alveolar, or velar) affected listeners' perceptual assimilation (Strange et al., 2001), but that this place of articulation affected discriminability and identification of English vowels differently (Nozawa, Frieda, & Wayland, 2003) and affected inexperienced learners more than experienced learners (Nozawa, Wayland, & Frieda, 2003). Assimilation of some British English vowels by Danish listeners (Bohn & Steinlen, 2003) and of French Parisian and North German vowels by American listeners (Strange, Levy, & Law, 2009) was also found to be strongly affected by consonantal context. Bohn and Steinlen (2003) asked participants to identify the eleven monophthongs of British English produced in three consonantal contexts: /hVt/, /dVt/, and /gVk/, and found that the perceptual assimilation of /ɪ, ɛ, ʊ, ʌ/ was strongly affected by consonantal context. For example, /ɪ/ was often assimilated as the Danish [e] in the /hVt/ and /dVt/ contexts but instead as the Danish [i] in the /gVk/ context. Conversely, the perceptual assimilation of /ɔː/ and /ɜː/ was affected little by the consonantal context, as the participants relied more on duration.

In summary, the results of the aforementioned studies show that examination of cross-linguistic perception of vowels cannot be reduced to only one phonetic context, but should rather include various consonantal contexts to better understand difficulties in L2 vowel learning.

1.5 French nasal vowels

1.5.1 Generality

Nasality is not an uncommon feature among languages. Many languages, such as French and English, have nasal consonants in their inventories. Out of the 451 languages listed in the UCLA Phonological Segment Inventory Database (UPSID), 78% possess an /n/, 76% have an

/m/, and 42% have an /ŋ/. Nasal vowels are also found in Bambara, Panjabi, Breton, and about forty other less commonly taught languages and dialects (Ruhlen, 1973), and among Indo-European languages, only French, Polish, and Portuguese possess nasal vowels (Straka, 1979)[2]. Specifically, the nasal vowels /ĩ/ and /ɑ̃/ are found in 13% of the languages, /ũ/ in 12%, /ɛ̃/ and /õ/ in 5%, /ɔ̃/ in 4%, and /ẽ/ in 2%. A language can possess the same number of oral and nasal vowels or can—like French—have a greater number of oral than nasal vowels. The contrary is, however, not attested in any language.

Another important distinction to make is the difference between nasal vowels and nasalized vowels, which are present in English in words such as "under". The nasal vowels are phonetically nasalized, but are also nasal from a phonological perspective in the sense that their nasality is what opposes them to oral vowels. Furthermore, a universal characteristic of nasal vowels is that their timbres differ from those of their oral counterparts. For instance, the close nasal vowels are more open than their oral counterparts. This is also the case with French nasal vowels.

1.5.2 Articulatory and acoustic characteristics

The particularity of nasal vowels is that they are produced with a lowering of the velum so that air escapes both through nose as well as the mouth, whereas oral vowels are produced with a closing of the velopharyngeal port. In the case of nasalized vowels, the velum is lowered, but this condition is not sufficient to let air go through the nasal cavity. In terms of aerodynamics, the resistance is twice as great in the nasal cavity as in the oral cavity, forcing the air to escape through the mouth. Previous studies have shown that to produce a nasal vowel or

[2] The realization of nasal vowels in Portuguese is different because a phonetic trace of the nasal consonant still exists (Galvao, 1998).

consonant, the velopharyngeal area has to be greater than 40 square centimeters (Warren, Dalston, & Mayo, 1993).

From an articulatory perspective, French nasal vowels differ in terms of height, as they do not have the same distance between the jaws, between the tongue and the palate, or between the upper and lower lips, the latter being the most visually salient cue. Another very salient difference between the three vowels is the rounding, as Figure 1 to Figure 3 show. Zerling (1989) suggests that the three nasal vowels have three distinct degrees of labiality, namely [-lab] or unrounded for [ɛ̃], [+lab] or rounded for [ɑ̃], and [++lab] or hyper-rounded for [ɔ̃][3]. The feature of rounding is particularly relevant for this study because, although English also has rounded vowels [i.e., u, ʊ, o, ɔ], the rounding in French is more marked and there is a strong protrusion of the lips. Therefore, English learners of French need to be able to visually determine the difference between rounded vowels in English and in French and to learn that French vowels can be much more rounded.

[3] Zerling (1989) also used the terms non labialized, labialized, and super labialized.

Figure 1. Articulatory characteristics of [ɔ̃] as produced in [gɔ̃k]

Figure 2. Articulatory characteristics of [ɑ̃] as produced in [gɑ̃g]

Figure 3. Articulatory characteristics of [ɛ̃] as produced in [gɛ̃k]

1.5.3 Variation and transcription

Historically, French used to distinguish four different nasal vowels [ɑ̃ - ɔ̃ - ɛ̃ - œ̃]. This distinction is still present in many dialects of French, such as Southern French (Durand, 1988; Walter, 1977), Southeastern French (Violin-Wigent, 2006), Québec French (Lappin, 1982; P. Léon, 1983; Martin, Beaudoin-Bégin, Goulet, & Roy, 2001), and to some measure in French spoken in Belgium (Pohl, 1983). However, since the middle of the twentieth century, [œ̃] has been progressively replaced by [ɛ̃] in Standard French and Parisian French so that now for the most part only three nasal vowels are used (Walter, 1994). Carton (1974) presents acoustical and articulatory reasons for the elimination of [œ̃]. Acoustically, the difference between [œ̃] and [ɛ̃] is difficult to perceive, and because nasalization neutralized rounding oppositions, "no place of [œ̃]" exists in the contrast between the three nasal vowels [ɑ̃ - ɔ̃ - ɛ̃]. Some consider that [œ̃] should no longer be considered as a phoneme (Battye, Hintze, & Rowlett, 2003), whereas others

note that the [œ̃- ɛ̃] contrast is not important anymore since context helps prevent confusion (Léon & Léon, 2007).

Interestingly, these three nasal vowels do not exactly correspond to the three oral vowels with which they share their International Phonetic Alphabet (IPA) symbol. Zerling (1989) showed that [ɑ̃] is articulated with a protrusion and a narrowing of the interlabial gap, making it a somewhat rounded vowel (as opposed to the unrounded [ɑ]) with labiality that is closer to the close-mid back rounded [ɔ]. The back nasal vowel [ɔ̃] is described as "hyper-rounded" and therefore more similar to the oral mid-close [o] than to [ɔ]. Finally, [ɛ̃] is an unrounded vowel often described as more open than [ɛ] with a degree of opening close to [æ] or [a] (Martinet, 1988). Despite the inadequacies mentioned above, the phonetic transcriptions of the stimuli used in this study will follow the IPA. Orthographically, the three nasal vowels can be represented by different graphemes so that [ɔ̃] is present in words with <on, om>, [ɛ̃] in words with <in, ain, ein, ien, ïn, en, yn, im>, and [ɑ̃] in words with <an, en, am, em, aon, aen>.

1.5.4 Previous L2 studies on the perception and production of French nasal vowels

Learners of French often encounter difficulty with the perception and production of French nasal vowels because their phonemic inventories do not possess nasal vowels. They often, however, possess vowels that are similar to French nasal vowels, but their nasalization occurs because of the phonetic environment (preceding a nasal consonant because of regressive assimilation) rather than as a distinction between minimal pairs. To date, studies have been conducted with learners of French with various L1s, such as Japanese (Racine, Detey, Schwab, & Zay, 2010; Takeuchi & Arai, 2009), Spanish (Racine et al., 2010), Brazilian Portuguese (Berri & Pagel, 2003), and American English (Inceoglu, 2011; Montagu, 2002). For example, Racine et al. (2010) conducted a corpus-based study on the production of nasal vowels by Japanese and

Spanish learners of French who participated in a repetition task and a reading aloud task. The authors based their analysis on a perceptive assessment by non-expert listeners (i.e., lexical identification task and confidence rating) and an acoustic analysis of the degree of postvocalic excrescence by linguists (i.e., the degree of presence of a postvocalic consonant). Relevant to the current study, the results of the lexical identification task showed that the rate of correctness was higher for Japanese learner productions (64.50%) than for Spanish learner productions (50.72%), that [ɔ̃] was better identified (67.02%) than [ɑ̃] (54.53%) and [ɛ̃] (51.27%), and that words produced in the reading task were better identified (60.42%) than those in the repetition task (54.78%). Results of the goodness of fit task showed a similar pattern, but closer observation of the data, although not discussed by the authors, seems to indicate that the production of [ɛ̃] by Japanese participants was actually better identified than the production of [ɑ̃], both in the repetition and reading tasks. A very low identification accuracy score for [ɛ̃] in the reading task by the Spanish learners might have skewed the results and presented a misleading representation of the pattern. The results of a small scale study conducted by Takeuchi and Arai (2009) also showed that Japanese learners of French do use lip protrusion when trying to produce [ɑ̃], indicating that the degree of lip protrusion is difficult to acquire for non-native speakers.

Brazilian Portuguese does possess nasal vowels (Mateus & D'Andrade, 2000), but the production of French nasal vowels by Portuguese Brazilian learners of French has been found to be problematic (Berri & Pagel, 2003). The productions of nine speakers were evaluated by native speaker phoneticians, and the results suggested that a wrong degree of labialization was the most common reason for non-native pronunciation. In addition, [ɑ̃] was found to be produced as the Portuguese vowel [ɐ̃], [ɛ̃] was often wrongly articulated as a back vowel, and [ɔ̃] appeared to be the least problematic vowel, probably because of the presence of an allophone in Portuguese.

As mentioned earlier, some French vowels are rounded and produced with a protrusion of the lips that does not exist in English. To understand labial adjustments made by learners of French, Montagu (2002) compared the productions of the two back nasal vowels [ɑ̃-ɔ̃] and two oral vowels [a-o] produced by eleven French and eleven American English speakers. Two video cameras (front and profile) recorded participants' productions of CV, CṼ, and CVN real words embedded in the sentence "Je répète ___ deux fois" (I repeat ___ twice). Based on measures of interlabial gap and lip protrusion—both expressed in cm^2 —Montagu reported that [ɑ̃] has a smaller interlabial gap but stronger protrusion than [a], whereas [ɔ̃] has an interlabial gap half the size of [o] but a similar degree of protrusion. What is particularly noteworthy is that the American English speakers failed to produce [ɔ̃] with a protrusion, accentuating the importance of directing the attention of learners to this non-native feature.

To get a better understanding of whether people are sensitive to visual cues, such as lip rounding, I investigated the perception of nasal vowels by 34 American English intermediate learners of French and 25 native French speakers (2011). The stimuli were presented in three conditions: A-only, AV, and V-only, and they consisted of 56 CV and four V words, where C was one of the following: /p-b-t-d-k-g-s-f-v-m-n-l-ʁ-ʒ/, and V was [ɛ̃], [õ], [ɑ̃], or the oral [o]. Participants were asked to watch and/or listen to the stimuli and circle the vowel they thought had been produced. Results showed that for the L2 learners, performance was better in the AV and A-only conditions and worse in the V-only condition, and better for the hyper-rounded [õ], followed by the unrounded [ɛ̃], and finally the rounded [ɑ̃]. However, different effects of modality were found across vowels. For instance, the identification performance for [ɛ̃] was the same across the different modalities, suggesting that participants could use both A and V information separately, but their integration into AV did not enhance accurate perception.

Furthermore, across all conditions, [ɑ̃] resulted in the most misidentification. In the AV modality, [ɑ̃] was perceived correctly 43% of the time but was perceived as [ɛ̃] 42% of the time. In the A modality, [ɑ̃] was perceived correctly 47% of the time, but was perceived as [ɛ̃] 36% of the time. However, participants identified [ɑ̃] correctly in the V condition only 23% of the time, probably due to its intermediate position on the continuum of labiality developed by Zerling (1989), which would indicate that it lacks visual saliency.

CHAPTER 2: CURRENT STUDY

2.1 Research questions and hypotheses

The current study was guided by the following research questions and associated hypotheses:

(1) Does AV perceptual training lead to greater improvement in L2 perception of French nasal vowels than A-only training does?

Previous literature on L1 speech perception suggests that participants benefit from visual information (Benoît et al., 1994). This has also been demonstrated in L2 speech perception of consonants, when the contrast is visually salient (e.g., Hardison, 1996, 2003, 2005b; Hazan et al., 2005; Wang et al., 2008, 2009), and of vowels (Hirata & Kelly, 2010; Soto-Faraco et al., 2007). In this study, participants in the AV training modality were predicted to perform better on the perception posttest than participants in the A-only training modality, because those in the AV modality were able to use visual information as an additional cue to discriminate the three French nasal vowels. This was predicted to be possible because the difference between the three nasal vowels is visually salient.

(2) Does AV perceptual training lead to greater improvement in L2 production of French nasal vowels than A-only training does?

Previous studies indicate that the effects of perception training can transfer to production skills (e.g., Bradlow et al., 1999; Hardison, 2003; Iverson et al., 2011; Wang, Jongman, & Sereno, 2003). The pronunciation of participants receiving AV training has also been shown to improve more than that of participants receiving A-only training (Hazan et al., 2005). Therefore perception training should improve production of French nasal vowels (as opposed to no training for the control group), and that improvement is expected to be greater for participants in the AV training group.

(3) Does perception accuracy vary in relation to consonantal context?

Previous studies have demonstrated that consonantal context affects perceptual assimilation of L2 French front rounded vowels (Levy, 2009a), perception of L2 French vowels (Gottfried, 1984), and vowel lipreading in English (Montgomery et al., 1987). Therefore, in the current study, participants' accurate perception of the vowels should be affected by the consonantal context despite the fact that coarticulatory effects of consonants on vowels may be restricted by the need to keep vowel phonemes distinct. In particular, the current study investigated the effect of twelve fricatives and occlusives with different places of articulation, and also looked at both word-initial consonants and word-final consonants. Initial labial consonants are predicted to possibly have more effects on the perception of following vowels because their articulatory characteristics might reduce the saliency of the vowel's visual cues.

(4) Is training generalizable to novel stimuli?

Previous studies have shown that perceptual training is generalizable to novel stimuli when either only one talker (Hardison, 2003; Motohashi-Saigo & Hardison, 2009) or several talkers (Hardison, 2003; Lively et al., 1993; Wang et al., 1999) were presented during training. The current study used only one talker during the training, and the results of the generalization test were predicted to be similar to those of the posttest.

2.2 Participants

The participants in this study were 60 American English learners of French (43 females, 17 males) ranging in age from 18 to 24 with an average age of 20 years (*SD* = 1.44). Forty participants were randomly assigned to one experimental group (AV or A perceptual training), and 20 served as controls (i.e., they participated in the pre/posttests but did not receive any training). The participants were intermediate-level learners of French enrolled in FRN202 at

Michigan State University during the Spring semester of 2013. Four sections were offered and taught by three different female teachers. With the instructors' and the program director's permissions, I visited the four classes and recruited participants who signed up for the study. I informed non-native speakers of English that they could not participate, but that other extra-credit opportunities were possible besides my study. On the first day of the study, I distributed a background questionnaire (Appendix A) to the participants which was used to control for possible additional L1s, stay-abroad influence (especially in Francophone Canada, southern France and other Francophone countries with different French dialects), amount of audiovisual input (e.g., movies, conversation with native speakers) and aural input (e.g., radio) in French, age, additional L2s, and background in French phonetics and linguistics (e.g., list of classes taken). The results of the questionnaires showed that none of the participants had stayed in a Francophone country for a period greater than two weeks, none of the participants had training in lipreading, and all reported good vision and no hearing disorders. One participant was enrolled as a linguistics major, but had just taken two general linguistics classes and had no previous knowledge of French phonetics. Participants reported watching videos in French 0.3 hours a week (*SD* = 0.81), listening to French music, radio or any other aural media 0.4 hours a week (*SD* = 0.79), and using French outside of the classroom 1.38 hours a week (*SD* = 1.38). Participants in the experimental group were paid $60 for their time and received 10 extra-credits in their French class. Participants in the control group were paid $20 and also received 10 extra-credits. The data of one participant from the control group were excluded from the analysis because she did not complete the tasks properly.

Two native speakers of French (including myself) served as raters. They were either trained in teaching French as a foreign language or in linguistics, and they were familiar with accented speech.

2.3 Material

Previous perception studies used either nonsense words (Iverson et al., 2011; Levy & Strange, 2008) or real words (Hardison, 2003). However, designing a discrimination study with triads of French nasal vowels makes use of only nonsense words or only real words impossible. Very often, for most of the triads, one stimulus/vowel would have to be excluded. Because the purpose of the experiment was to identify sounds, the task for the listeners was to focus only on the specific sounds under test, and learners' potential background vocabulary knowledge was unlikely to influence their identification of the perceived vowels. In addition, previous studies have demonstrated that no correlation exists between word familiarity and identification scores in speech perception experiments (e.g., Flege, Takagi, & Mann, 1995). Therefore, for the current study, a mix of real words and nonsense words was used.

Stimuli were based on triads of the three French Parisian vowels [ɑ̃ - ɔ̃ - ɛ̃] in various consonantal contexts. A total of 396 stimuli were recorded for either the pretest and posttest, the training, or the generalization test. None of the tokens was presented during both testing and training. The stimuli had the following structures:

a) A total of 324 #CVC#, where C was one of the following consonants [p-t-k-b-d-g-s-z-f-v-ʒ-ʃ]. For example: [pɔ̃t].
b) A total of 36 initial consonant clusters #CcVC#, where the cluster was [dʁ] and the final consonant was one of [p-t-k-b-d-g-s-z-f-v-ʒ-ʃ]. For example: [dʁɔ̃t].

c) A total of 36 final consonant clusters #CVCc#, where the first consonant was one of the following [p-t-k-b-d-g-s-z-f-v-ʒ-ʃ] and the final cluster was [dʁ]. For example: [kɔ̃dʁ].

The list of pretest stimuli was comprised of 108 items with a balanced distribution of vowels and of placement and manner of articulation for the consonants. The distribution of initial and final consonants for [ɔ̃], [ɑ̃], and [ɛ̃] is provided in Table 1, Table 2, and Table 3, respectively. There was a total of 36 tokens per vowel, with initial and final consonants equally distributed between manner (occlusive and fricative) and place (bilabial, labiodental, dental, alveolar, palatal, and velar) of articulation, thus resulting in six tokens per manner/place of articulation. Note that the exact same initial and final consonants were not used for the three vowels, but rather the voiced or voiceless counterparts (i.e., with the same manner and place of articulation). For instance, the initial bilabial/final bilabial token for [ɔ̃] was [pɔ̃p], but was [bɑ̃p] for [ɑ̃], and [bɛ̃b] for [ɛ̃]. The rationale for not using the same consonants was to prevent participants from directly contrasting stimuli if two appeared consecutively in a trial, and to prevent them from memorizing stimuli they had already heard—although this was less plausible due to the large number of stimuli.

Table 1.

Distribution of [ð] for the pretest and posttest (n = 36)

				Final consonant											
				Occlusive						Fricative					
				Bilabial		Dental		Velar		Labio-dental		Alveolar		Palatal	
				[p]	[b]	[t]	[d]	[k]	[g]	[f]	[v]	[s]	[z]	[ʃ]	[ʒ]
Initial consonant	Occlusive	Bilabial	[p]	[ð]					[ð]	[ð]					
			[b]					[ð]			[ð]	[ð]			
		Dental	[t]		[ð]				[ð]						[ð]
			[d]			[ð]							[ð]	[ð]	
		Velar	[k]	[ð]		[ð]					[ð]		[ð]		
			[g]				[ð]								[ð]
	Fricative	Labio-dental	[f]					[ð]					[ð]	[ð]	
			[v]		[ð]				[ð]	[ð]					
		Alveolar	[s]			[ð]						[ð]		[ð]	
			[z]	[ð]			[ð]								[ð]
		Palatal	[ʃ]					[ð]			[ð]				
			[ʒ]		[ð]		[ð]			[ð]		[ð]			

Table 2.

Distribution of [ã] for the pretest and posttest (n = 36)

				Final consonant											
				Occlusive						Fricative					
				Bilabial		Dental		Velar		Labio-dental		Alveolar		Palatal	
				[p]	[b]	[t]	[d]	[k]	[g]	[f]	[v]	[s]	[z]	[ʃ]	[ʒ]
Initial consonant	Occlusive	Bilabial	[p]			[ã]					[ã]		[ã]		
			[b]	[ã]			[ã]							[ã]	
		Dental	[t]					[ã]		[ã]					
			[d]		[ã]				[ã]		[ã]	[ã]			
		Velar	[k]		[ã]										[ã]
			[g]			[ã]		[ã]				[ã]		[ã]	
	Fricative	Labio-dental	[f]			[ã]									[ã]
			[v]	[ã]			[ã]						[ã]	[ã]	
		Alveolar	[s]				[ã]			[ã]					
			[z]		[ã]				[ã]		[ã]	[ã]			
		Palatal	[ʃ]	[ã]					[ã]	[ã]					[ã]
			[ʒ]					[ã]					[ã]		

Table 3.

Distribution of [ɛ̃] for the pretest and posttest (n = 36)

				Final consonant											
				Occlusive						Fricative					
				Bilabial		Dental		Velar		Labio-dental		Alveolar		Palatal	
				[p]	[b]	[t]	[d]	[k]	[g]	[f]	[v]	[s]	[z]	[ʃ]	[ʒ]
Initial consonant	Occlusive	Bilabial	[p]				[ɛ̃]	[ɛ̃]					[ɛ̃]	[ɛ̃]	
			[b]		[ɛ̃]										[ɛ̃]
		Dental	[t]	[ɛ̃]		[ɛ̃]							[ɛ̃]	[ɛ̃]	
			[d]				[ɛ̃]			[ɛ̃]					
		Velar	[k]					[ɛ̃]		[ɛ̃]		[ɛ̃]			
			[g]	[ɛ̃]					[ɛ̃]		[ɛ̃]				
		Labio-dental	[f]		[ɛ̃]				[ɛ̃]	[ɛ̃]		[ɛ̃]			
			[v]			[ɛ̃]					[ɛ̃]				
	Fricative	Alveolar	[s]		[ɛ̃]				[ɛ̃]		[ɛ̃]				[ɛ̃]
			[z]					[ɛ̃]					[ɛ̃]		
		Palatal	[ʃ]				[ɛ̃]					[ɛ̃]		[ɛ̃]	
			[ʒ]	[ɛ̃]		[ɛ̃]									[ɛ̃]

Training stimuli originally consisted of 180 CVC stimuli with various phonetic contexts, but two tokens (i.e., [gɔ̃ʒ] and [bɑ̃z]) had to be removed because of poor audio quality, thus reducing the total number of tokens to 178. The distribution of initial and final consonants for the training of [ɔ̃], [ɑ̃], and [ɛ̃] is provided in Table 4, Table 5, and Table 6, respectively. There was a total of 59 tokens for [ɔ̃] and [ɑ̃], and 60 tokens for [ɛ̃]. Similar to the pre/posttest stimuli, the initial and final consonants of the training stimuli were equally distributed between manner

(occlusive and fricative) and place (bilabial, labiodental, dental, alveolar, palatal, and velar) of articulation, thus resulting in 10 tokens per manner/place of articulation. Again, note that the exact same initial and final consonants were not used for the three vowels, but rather the voiced or voiceless counterparts (i.e, with the same manner and place of articulation).

Table 4.

Distribution of [ɔ̃] for the training (n = 59)

				Final consonant											
				Occlusive						Fricative					
				Bilabial		Dental		Velar		Labio-dental		Alveolar		Palatal	
				[p]	[b]	[t]	[d]	[k]	[g]	[f]	[v]	[s]	[z]	[ʃ]	[ʒ]
Initial consonant	Occlusive	Bilabial	[p]			[ɔ̃]		[ɔ̃]			[ɔ̃]	[ɔ̃]		[ɔ̃]	
			[b]		[ɔ̃]		[ɔ̃]		[ɔ̃]	[ɔ̃]					[ɔ̃]
		Dental	[t]	[ɔ̃]		[ɔ̃]		[ɔ̃]			[ɔ̃]		[ɔ̃]		
			[d]		[ɔ̃]		[ɔ̃]			[ɔ̃]		[ɔ̃]			[ɔ̃]
		Velar	[k]		[ɔ̃]		[ɔ̃]		[ɔ̃]			[ɔ̃]			
			[g]	[ɔ̃]				[ɔ̃]			[ɔ̃]		[ɔ̃]	[ɔ̃]	
		Labio-dental	[f]		[ɔ̃]		[ɔ̃]			[ɔ̃]		[ɔ̃]			[ɔ̃]
			[v]	[ɔ̃]		[ɔ̃]		[ɔ̃]			[ɔ̃]		[ɔ̃]		
	Fricative	Alveolar	[s]	[ɔ̃]			[ɔ̃]		[ɔ̃]	[ɔ̃]					[ɔ̃]
			[z]			[ɔ̃]		[ɔ̃]			[ɔ̃]		[ɔ̃]	[ɔ̃]	
		Palatal	[ʃ]		[ɔ̃]		[ɔ̃]		[ɔ̃]			[ɔ̃]			[ɔ̃]
			[ʒ]	[ɔ̃]				[ɔ̃]			[ɔ̃]		[ɔ̃]	[ɔ̃]	

Table 5.

Distribution of [ã] for the training (n = 59)

				Final consonant											
				Occlusive						Fricative					
				Bilabial		Dental		Velar		Labio-dental		Alveolar		Palatal	
				[p]	[b]	[t]	[d]	[k]	[g]	[f]	[v]	[s]	[z]	[ʃ]	[ʒ]
Initial consonant	Occlusive	Bilabial	[p]	[ã]			[ã]		[ã]	[ã]				[ã]	
			[b]			[ã]		[ã]			[ã]				[ã]
		Dental	[t]		[ã]		[ã]				[ã]	[ã]			
			[d]	[ã]		[ã]		[ã]		[ã]			[ã]	[ã]	
		Velar	[k]	[ã]				[ã]		[ã]		[ã]		[ã]	
			[g]		[ã]		[ã]		[ã]				[ã]		[ã]
		Labio-dental	[f]	[ã]			[ã]				[ã]		[ã]		
			[v]		[ã]	[ã]			[ã]	[ã]		[ã]			[ã]
	Fricative	Alveolar	[s]		[ã]	[ã]			[ã]		[ã]	[ã]		[ã]	
			[z]				[ã]	[ã]		[ã]					[ã]
		Palatal	[ʃ]		[ã]			[ã]			[ã]		[ã]	[ã]	
			[ʒ]	[ã]		[ã]			[ã]			[ã]			[ã]

Table 6.

Distribution of [ɛ̃] for the training (n = 60)

				Final consonant											
				Occlusive						Fricative					
				Bilabial		Dental		Velar		Labio-dental		Alveolar		Palatal	
				[p]	[b]	[t]	[d]	[k]	[g]	[f]	[v]	[s]	[z]	[ʃ]	[ʒ]
Initial consonant	Occlusive	Bilabial	[p]		[ɛ̃]	[ɛ̃]			[ɛ̃]	[ɛ̃]					[ɛ̃]
			[b]				[ɛ̃]	[ɛ̃]			[ɛ̃]		[ɛ̃]	[ɛ̃]	
		Dental	[t]		[ɛ̃]		[ɛ̃]			[ɛ̃]		[ɛ̃]			[ɛ̃]
			[d]	[ɛ̃]		[ɛ̃]			[ɛ̃]		[ɛ̃]		[ɛ̃]		
		Velar	[k]	[ɛ̃]		[ɛ̃]			[ɛ̃]		[ɛ̃]		[ɛ̃]	[ɛ̃]	
			[g]		[ɛ̃]			[ɛ̃]				[ɛ̃]			[ɛ̃]
		Labio-dental	[f]	[ɛ̃]		[ɛ̃]		[ɛ̃]			[ɛ̃]		[ɛ̃]		
			[v]		[ɛ̃]		[ɛ̃]			[ɛ̃]		[ɛ̃]		[ɛ̃]	
	Fricative	Alveolar	[s]			[ɛ̃]		[ɛ̃]		[ɛ̃]			[ɛ̃]	[ɛ̃]	
			[z]		[ɛ̃]		[ɛ̃]		[ɛ̃]		[ɛ̃]				[ɛ̃]
		Palatal	[ʃ]	[ɛ̃]				[ɛ̃]					[ɛ̃]		[ɛ̃]
			[ʒ]		[ɛ̃]		[ɛ̃]		[ɛ̃]	[ɛ̃]		[ɛ̃]		[ɛ̃]		

The generalization stimuli list included 36 novel CVC tokens similar to those of the pre/posttest and training, with [ɔ̃], [ɑ̃], and [ɛ̃] in various consonantal contexts. To examine whether training was also generalizable to novel consonantal contexts, 36 tokens were presented with [dʁ] as the initial consonantal cluster and one of the following [p-t-k-b-d-g-s-z-f-v-ʒ-ʃ] as the final consonant. Conversely, 36 tokens were presented with [dʁ] as a final consonant cluster and one of the above consonants as the initial consonant. The reason for including the consonant cluster [dʁ] for the novel stimuli was that the uvular [ʁ], which was not one of the consonants

during training and pre/posttest, does not exist in English. Combined with the plosive dental [d], they form a sound that is novel and that could possibly be challenging for L2 listeners. Refer to Appendix B for the list of stimuli for the generalization test. The total duration of the posttest session was about one hour.

2.4 Recording

A female native speaker of French in her early thirties was video-recorded in a quiet research room at Michigan State University. She spoke the standard variety of French and did not make differences between the vowels [œ̃] and [ɛ̃] during either unscripted speech (e.g., conversations with me) or during scripted speech (e.g., read sentences such as "un bon vin blanc" [œ̃ bɔ̃ vɛ̃ blɑ̃]). On the day of the recording, she was instructed to read a list of real and nonsense words containing nasal vowels in the most natural fashion with the aim of avoiding hyper-articulation. Each list ended with the repetition of the last utterance to control for list-final intonation. The camera captured a full-sized image of the speaker's head and her lower jaw drop was fully visible (see Figure 4). The recording was used for both the AV and the A modalities to prevent any possible aural variations for each token across the two modalities. Each utterance was saved as one file to permit random ordering for the three testing modalities (AV, A and V) and for the two types of training modalities (AV or A).

Figure 4. Screen capture of the speaker's face

To assess the intelligibility of the stimuli, three monolingual native speakers of (Parisian) French who did not reside in the United States were asked to identify each token by indicating which nasal vowel –<on>, <un>, or <an>– was produced. They made 0 errors in identification, so none of the stimuli were re-recorded.

2.5 Procedure

For the pretest, posttest, and test of generalization, participants met individually with me in a quiet research room at Michigan State University. Participants were asked to come on the first day for the pretest (production and perception) and were invited to read the consent form and fill out a language background questionnaire. Scheduling for the upcoming training sessions was then arranged. Training sessions were carried out in the same computer lab, but no more than 6 participants were present in the lab at the same time in order for the researcher to be able

to attend to potential technical issues and to ensure that participants were completing the tasks appropriately. A summary of the procedure is displayed in Table 7.

Table 7.

Summary of the data collection procedure

Time 1	Consent Form Language Background Questionnaire Production pretest (video-recorded) (108 stimuli) Perception pretest (108 stimuli x 3 modalities (AV-A-V) = 324 stimuli)	Duration: about 1 hour
Time 2	Training session 1 (178 stimuli per session)	Duration: about 30 minutes
Time 3	Training session 2	
Time 4	Training session 3	
Time 5	Training session 4	AV or A training
Time 6	Training session 5	
Time 7	Training session 6	
Time 8	Production posttest (same 108 stimuli as pretest) Perception posttest (same as pretest, re-randomized) Perception generalization test	Duration: about 1 hour

2.5.1 Pretest

All the participants (trainees and controls) were administered two pretests: a perception and a production pretest. All listeners were tested during the month of March 2013. In the perception pretest, participants were presented with 108 stimuli in each of the three modalities: AV, A, and visual only (V), yielding a total of 324 responses per participant. The number of stimuli was not higher to ensure that participants would not get tired and to keep the duration of

the whole pretest session under one hour. The V modality was included for testing to shed some light on whether visual cues also play a significant role in vowel perception and recognition when audio stimuli are absent.

Before the beginning of the pretest, I informed the participants that the study was about the three nasal vowels [ɔ̃], [ɑ̃] and [ɛ̃]. I told them that [ɔ̃] was the same sound as in commonly known words such as “pont” (bridge), “onze” (eleven), “poisson” (fish), “maison” (house) and “blond” (blond) and was often represented by the letters <on>, the sound [ɑ̃] was the one in words such as “cent” (hundred), “dent” (teeth), “vent” (wind), “plan” (map) and “blanc” (white) and could be written with the letters <an> like the word for year, and finally the sound [ɛ̃] was present in words like “main” (hand), “pain” (bread), “train” (train), “vingt” (twenty) and “chien” (dog), and could be written <un> like the number one. I then informed them that the sound [ɔ̃] would always be the option on the left and would be written <on>, the sound [ɑ̃] would be in the middle and written <an>, and the sound [ɛ̃] would be the option on the right and would be written <un>. The order of the tokens heard was different for each modality and was counterbalanced across participants. Two orders were used for the presentation of the three modalities: (AV, A, V) or (A, AV, V), and the two orders were also counterbalanced across participants. Participants sat in front of an iMac computer equipped with a 21.5-inch screen, running at 1920x1080, and the aural stimuli (in the A and AV modalities) were presented via high quality headphones. For each modality of the presentations, the experiments started with four practice stimuli to ensure that the participants were familiar with the task, that the volume was adequate (for the A and AV modalities), and that they were responding within four seconds. The four practice tokens did not contain any nasal vowels and were as follows: [ʃa], [vit], [sis] and [ʒon]. Participants were asked to choose between the sounds [a], [i] and [o]. Right after the

practice task, a message appeared on the screen prompting participants to click on “commencer” to start the experiment. Immediately after hearing a stimulus containing one of the three nasal vowels, the three choices appeared on the computer screen, and participants had to click on the correct option (see Figure 5).

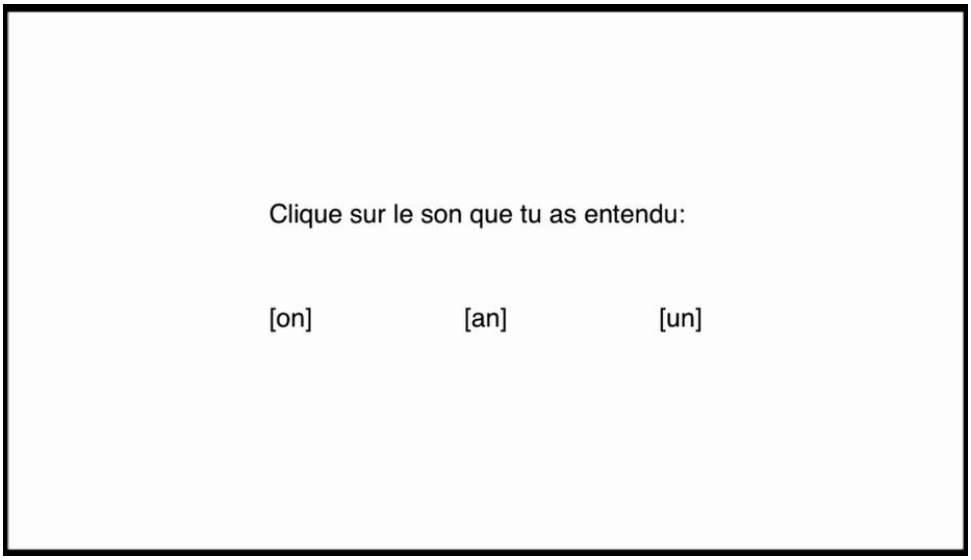

Figure 5. Prompt for the participants to select an answer

The stimulus, either AV, A or V, was played before the three options appeared on the screen so as not to restrict participants’ initial recognition of the stimulus and limit priming effects from the other options. In order to reduce the risk of confusion, [ɔ̃] options always appeared on the left, [ɑ̃] options in the middle, and [ɛ̃] options on the right. For example, participants heard [pɛ̃t] and were asked whether the vowel presented sounded like <on>, <an>, or <un>. Participants had 4 seconds to click on one of the three options before being presented with the next stimulus. They were encouraged to guess for stimuli that they were unsure of, and stimuli that were not identified were classified as incorrect. No feedback was given. The perception pretest lasted about 30 minutes, with ten minutes per modality.

The production pretest was carried out before the perception test to avoid influence from hearing the stimuli since the stimuli were the same in both experiments. Participants were presented with a spoken stimulus followed by one of the following sentences prompting them to repeat the word: “répète le mot s’il te plait”, “à ton tour de répéter”, “s’il te plait, répète le mot”. No orthographic support was presented. The reason for having a delay between the stimulus and its repetition, combined with the intervening speech material (i.e., the prompts), was to prevent direct imitation from sensory memory. The task was video-recorded with the webcam at the top of the computers, thus capturing full images of the participants’ faces. The program iMovie was used for the recording, but the window was closed so that the participants could not see themselves during the recording. The task lasted 9 minutes with a break in the middle and started with a practice task (i.e., no nasal vowels) to ensure that participants had understood the procedures.

2.5.2 Training

Based on their pretest scores, each participant was assigned to one training group: AV or A in order to ensure that groups were as balanced as possible. Each group heard the exact same stimuli and followed the same procedure, with the only difference being that the A group did not see the speaker’s face. All the training sessions followed the same procedures as for the pretest, except that participants received feedback. If the answer they clicked on was correct, the message “Bravo!” appeared in green, as illustrated on Figure 6.

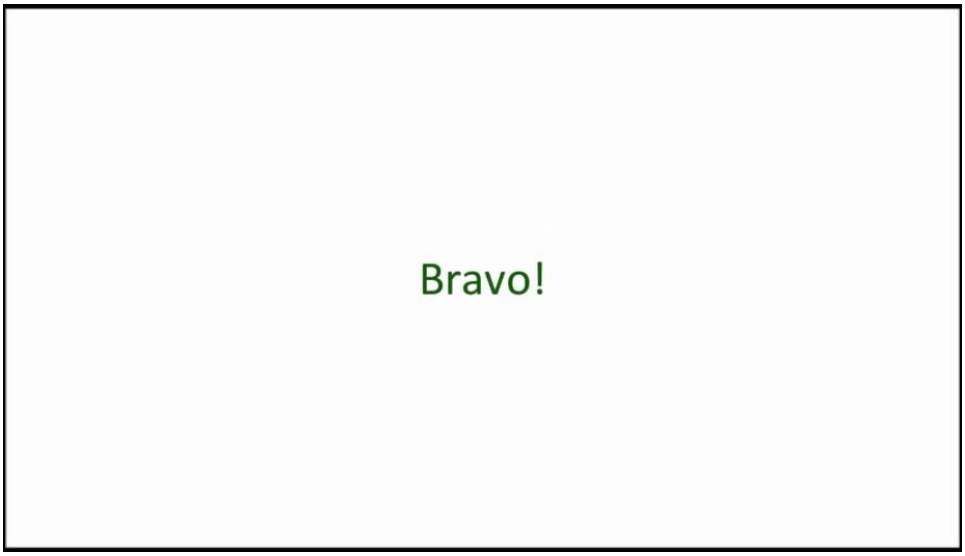

Figure 6. Message following a correct response

If the answer was incorrect, the message “incorrect” appeared in red, followed by the correct option. For instance, if a participant was presented with the stimulus [tɑ̃f] but clicked on <on>, the message as shown in Figure 7 would appear.

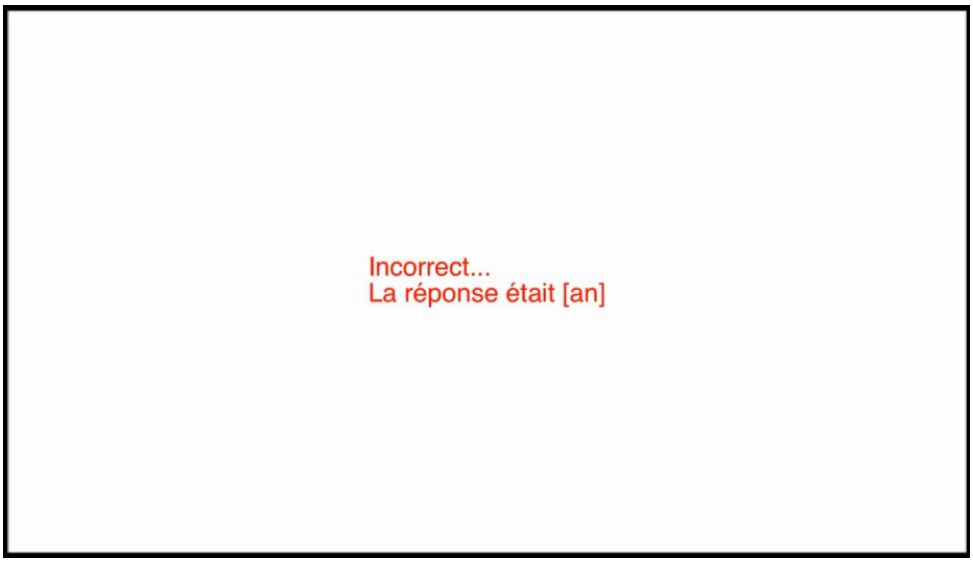

Figure 7. Message following an incorrect response

Participants in the AV training group were reminded that they had to watch the screen and should not close their eyes to concentrate. Because the entire experiment, including the

response task, was conducted on the computer, the chance for participants in the AV training group to not look at the visual stimuli was highly reduced. Training sessions consisted of 178 stimuli and lasted about 30 minutes with short breaks after each block of 30 stimuli. All participants completed their six training sessions within a period of 18 days (mean = 15 days).

2.5.3 Posttest and generalization test

After the last training session for the trained participants or after a period of 14 to 18 days for the control group, each participant took the posttest. They were asked to produce the same 108 stimuli as on the pretest to assess improvement in production and thereby the potential effects of the training. The procedure was the same as for the pretest. Participants then completed the perception posttest, which was comprised of the same stimuli as for the pretest but re-randomized. Following the posttest, a generalization test with novel stimuli was administered. The generalization test consisted of presentation of novel stimuli in the three modalities: AV-A-V or A-AV-V.

2.6 Perceptual rating

Two native speakers of Parisian French familiar with accented speech but with no formal background in phonetics judged the tokens produced by the learners in two perceptual evaluation tests. They were paid 150 euros for their participation in the study. The first test consisted of a minimal-pairs identification task. Raters were asked to focus on the realizations of the nasal vowel rather than on the pronunciation of the consonants or the word itself. Tokens were presented in a three-alternative forced choice task using TP software designed for speech perception experiments (Rauber, Rato, Kluge, & Santos, 2012). Upon listening to the token, raters were presented with three options on the screen: “on”, “an”, and “un”, and they indicated

their choice by clicking on one of the three options. The task was not timed, and raters had the option to click a replay button to listen to a segment a second time.

The second test was a quality rating task. Raters were presented with the intended word on a computer screen and were asked to rate the L2 learners' production of the vowels on a scale from 1 (bad) to 7 (excellent). They were asked to ignore the production of the initial and final consonants. In order to prevent fatigue and risk of confusion from continuously switching between the three vowels, the 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. For instance, all the tokens of [dɛ̃f] were grouped in one single block and the raters were asked to evaluate the quality of the production without knowing whether they were listening to a pretest or posttest production. The order of the stimuli produced by the L2 speakers was randomized within each block, and the presentation order of the different blocks was counterbalanced across raters. Raters took about 4 minutes to rate each block, and no more than ten blocks were presented in one session. The two rating tasks were presented via headphones at a comfortable listening level.

2.7 Analysis

All the statistical analyses were run using the SPSS software (version 19). To investigate training effects on the perception and production of the French nasal vowels, I compared the scores obtained at the perception tests and oral production tests in the pretest, posttest, and generalization posttest using repeated-measures ANOVAs. The mean correct identification score for each experimental group was tabulated and separate ANOVAs were conducted on pretest and posttest identification scores for each vowel and each modality of presentation.

To investigate the effect of perception training on production, the vowels produced by all the L2 participants (trainees and controls) in the pretest and posttests were presented to two

native French speakers, and identification accuracy scores and quality rating scores were obtained. The agreement rate between the two raters at the identification task was 88.4%, which is considered an acceptable reliability. Because of the high number of production tokens that raters were initially asked to judge (12,960), discrepancies were not further discussed. Instead, I rated the 1,540 tokens that raters disagreed on and used the response that one of the raters and I agreed on for the analysis. This way, all the responses were agreed upon by two native raters. For the quality rating task (7-point Likert scale), I calculated the average between the rating of the two French raters and used it for the analysis.

When using ANOVAs, there are certain assumptions that need to be checked. These assumptions are that (1) the data are normally distributed, (2) the variance in each group is similar, (3) the data are measured at the interval level, and (4) the data are independent. In this study, the fourth assumption was satisfied as no participant's score affected another participant's score and each participant was associated with only one experimental group. I ran a test of homogeneity of variance for each of the dependent variables which showed no violation of this assumption (i.e., $p > .05$ on the Levene's test of equality of variance). The third assumption was also met because the dependent variables were not dichotomous or categorical. Next, the results of a Kolmogorov-Smirnov test indicated that the data were normally distributed, except for the perception pretest in the AV-only testing modality. The results of the analysis should, therefore, be interpreted with caution. Finally, in repeated-measures designs, the assumption of sphericity must also be checked to verify if the differences between the variances of a single participant's data are equal (Larson-Hall, 2010, p. 336). When analyzing the data, I used the Mauchly's sphericity test to check whether the assumption of sphericity was violated, and if it was violated, I reported the Greenhouse-Geisser adjusted scores.

CHAPTER 3: RESULTS

The results presented in this chapter are organized by research questions. I first look at the comparability of the groups at the pretest, before examining the effects of training on the perception tests according to the modalities and the vowels. I then turn to the results of the effects of perceptual training on production. Then, I examine the effects of the consonantal context on the perception tests, and finally the results for the generalization test. For the statistical analysis, the alpha level was set at .05 ($\alpha = .05$).

3.1 Research question 1: Perception

None of the participants scored higher than 78% on the pretest. The risk of a ceiling effect was avoided as all participants had room to improve. No data were excluded from the analysis apart from one participant from the control group who did not complete the task correctly. The data presented in this chapter therefore stemmed from 59 participants divided into three groups: 20 AV training, 20 A-only training, and 19 control.

3.1.1 Comparability of the groups at the pretest

The grand mean accuracy scores[4] at the perception pretest were 42% (*SD* = 16.6) for the AV group, 44% (*SD* = 14.4) for the A group, and 47% (*SD* = 13.8) for the control group. In order to examine whether the three groups were statistically similar at the pretest, a one-way ANOVA was performed with Group (A, AV, V) as the independent variable and Accuracy score as the dependent variable. Results confirmed that the three groups were equal at pretest, $F(2, 56) = 0.45$, $p = .640$.

Further analyses were conducted in order to assess the comparability of the three groups in regards to the vowels and the presentation modalities. A repeated-measures ANOVA with

[4] The grand mean accuracy scores refer to the means across the three modalities of presentation (AV, A, and V) and the three French nasal vowels.

Vowel ([ɑ̃ - ɔ̃ - ɛ̃]) and Modality (A, AV, V) as within-subjects variables, and Training Group as a between-subjects variable was performed. Results indicated that the main effect of Vowel, $F(2, 112) = 19.11$, $p < .001$, and Modality, $F(2, 112) = 165.32$, $p < .001$, were significant, but that the interactions Group × Vowel and Group × Modality were not significant, $F_{G\times V}(4, 112) = 0.29$, $p = .879$ and $F_{G\times M}(4, 112) = 0.82$, $p = .514$, indicating that the three groups performed comparably in regards to vowels and modalities. Table 8 shows the accuracy scores at pretest averaged over the three treatment groups (AV training, A training, and control) for the three vowels and the three modalities of presentation. Post-hoc pairwise comparisons revealed statistically significant differences between each vowel within each of the three modalities, between AV and V for all vowels, between A and V for [ɔ̃] and [ɑ̃], and between A and AV for [ɑ̃].

Table 8.

Accuracy scores at pretest averaged over treatment groups

	[ɔ̃]	[ɑ̃]	[ɛ̃]
AV modality	73	38	27
A modality	72	45	29
V modality	58	26	32

The results of the pretest illustrated in Figure 8 show that two different hierarchies of identification were found. In the AV modality, L2 learners in the three groups were better at correctly identifying the vowel [ɔ̃] with a combined average of 73%, followed by [ɑ̃] (38%) and [ɛ̃] (27%). The results of the A modality followed the same patterns: [ɔ̃] (72%), [ɑ̃] (45%), and [ɛ̃] (29%). In the V modalities, [ɔ̃] was still the most discernible vowel (58%), but [ɛ̃] (32%) was found more intelligible than [ɑ̃] (26%). Post-hoc LSD tests showed that scores for each vowel

differed significantly from one another in the three modalities (mostly $p < .001$ and $p = .007$ for the difference between [ɛ̃] and [ɑ̃] in the V modality).

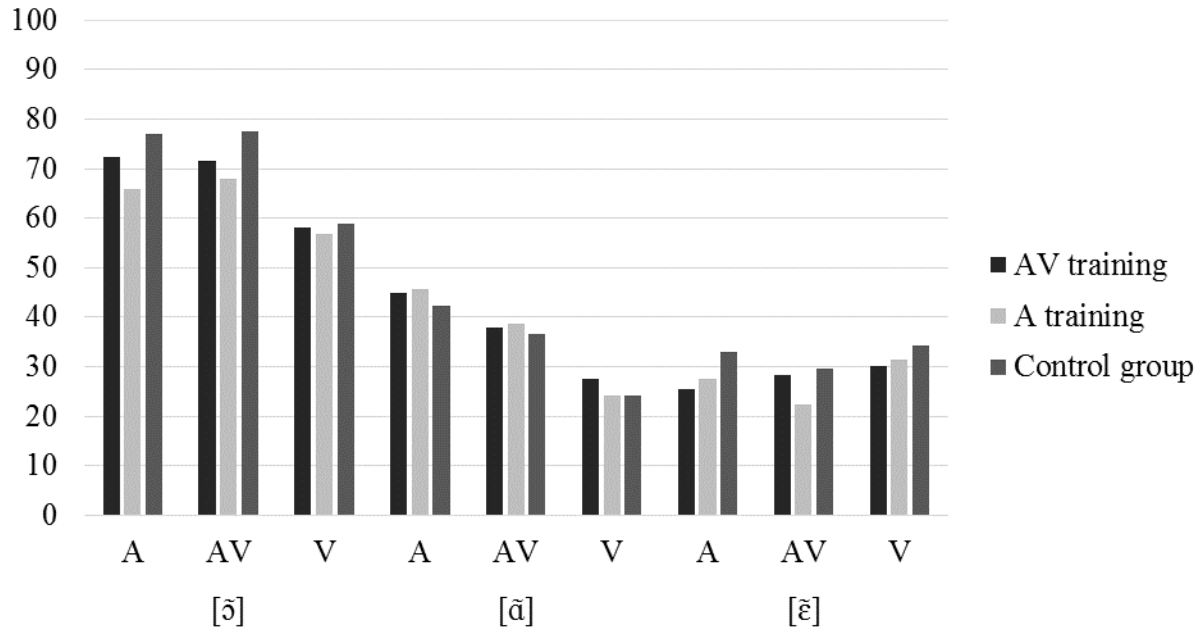

Figure 8. Percentage of correct identification score at the pretest for each vowel and modality

The results of perception confusion among the three nasal vowels (e.g., a participant thought s/he saw/heard [ɔ̃] when [ɑ̃] was actually produced) are displayed in the confusion matrix in Table 9. Across the three modalities, the vowel [ɔ̃] was accurately perceived more than half the time, and—in instances of confusion—was more often misperceived as [ɛ̃] than it was as [ɑ̃]. In the Audio modality, [ɑ̃] was accurately perceived less than half the time (45%), and was more often misperceived as [ɔ̃] (about 37%) than as [ɛ̃] (about 18%). When the visual component was added, the percentage of incorrect perception of [ɑ̃] as [ɔ̃] increased further (43%). The pattern was even stronger in the V modality where [ɑ̃] was twice as often perceived incorrectly as [ɔ̃] (54%) than as [ɑ̃] (26%), and also often perceived as [ɛ̃] (about 20%). Finally, [ɛ̃] was

rarely misperceived as [ɔ̃] across the three modalities, but was overwhelmingly misperceived as [ɑ̃] in the A (64%), AV (68%), and V (60%) modalities.

Table 9.

Confusion matrix for vowel identification at pretest (mean percent response)

		Experimental group								
		[ɔ̃]			[ɑ̃]			[ɛ̃]		
		AV	A	C	AV	A	C	AV	A	C
	Perceived as (in %)									
A	[ɔ̃]	**73**	**66**	**79**	37	39	36	7	8	5
	[ɑ̃]	10	10	7	**45**	**46**	**44**	67	65	61
	[ɛ̃]	17	23	14	18	15	20	**26**	**28**	**35**
AV	[ɔ̃]	**72**	**68**	**79**	44	45	39	5	6	4
	[ɑ̃]	8	7	6	**38**	**39**	**38**	66	71	66
	[ɛ̃]	20	25	15	18	17	23	**28**	**22**	**30**
V	[ɔ̃]	**58**	**57**	**59**	57	56	49	8	7	6
	[ɑ̃]	11	11	14	**28**	**24**	**25**	62	61	58
	[ɛ̃]	31	32	27	16	20	26	**30**	**31**	**36**

Note. Accurate perceptions appear in bold.

3.1.2 Analysis of the effectiveness of the training within groups

To assess the effectiveness of training condition, separate repeated-measures ANOVAs were conducted on the pretest and posttest identification scores for the two training groups. The variables were Time (pretest, posttest), Modality (A, AV, V) and Vowel (ɔ̃, ɑ̃, ɛ̃). Table 10 summarizes the scores for the two training groups and the control group at the pretest and posttest according to the modality of presentation. The results of the control group were not

significantly different from the pretest to the posttest in the A modality [$F(1, 18) = 0.29$, $p = .590$, $\eta_p^2 = .016$], the AV modality [$F(1, 18) = 2.47$, $p = .13$, $\eta_p^2 = .121$], and the V modality [$F(1, 18) = 0.98$, $p = .33$, $\eta_p^2 = .052$], indicating that any change in performance for the training groups was due to the training they received.

Table 10.

Accuracy scores for the perception pretest and posttest per group

	Modality	AV training group	A-only training group	Control group
Pretest	A	47.5% (16.1)	46.3% (15.8)	52.0% (15.4)
	AV	45.8% (18.4)	42.8% (14.1)	48.5% (15.8)
	V	38.5% (21.7)	37.4% (17.2)	39.6% (15.2)
Posttest	A	76.8% (18.4)	74.6% (18.1)	53.4% (18.7)
	AV	79.5% (18.3)	73.1% (18.2)	53.1% (17.7)
	V	63.3% (18.2)	60.8% (18.5)	41.5% (15.6)

Note. Standard deviations are in parentheses.

For the AV training group, there was a significant main effect of Time, $F(1, 19) = 55.56$, $p < .001$, $\eta_p^2 = .74$, Modality, $F(2, 38) = 18.38$, $p < .001$, $\eta_p^2 = .49$, and Vowel, $F(2, 38) = 40.75$, $p < .001$, $\eta_p^2 = .68$. The interaction of Time × Modality, $F(2, 38) = 2.84$, $p = .07$, $\eta_p^2 = .13$ indicates that perception gains over time were comparable for the three modalities. Comparisons of posttest performances (Figure 9) showed an increase of about 29% in the A modality, 34% in the AV modality and 25% in the V modality. The interaction Time × Vowel, however, was significant $F(2, 38) = 12.42$, $p < .0001$, $\eta_p^2 = .39$ with a performance increase of 44% for [ɛ̃], 23% for [ɑ̃], and 20% for [ɔ̃]. Also, the interaction Time × Modality × Vowel was significant, $F(4, 76) = 5.04$, $p = .001$, $\eta_p^2 = .21$. The p values reported in Table 11 show that the correct

identification scores for each vowel were significantly different from one another across the three modalities, indicating that the perception of some vowels improved significantly more than others.

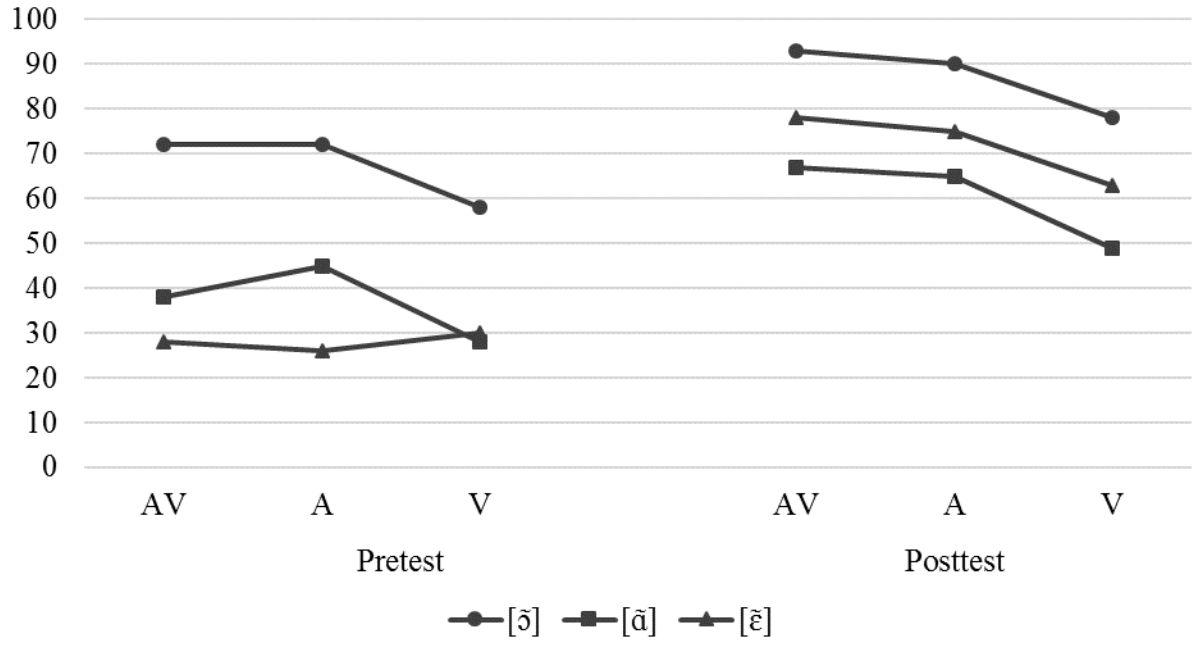

Figure 9. Changes in perceptual accuracy of French nasal vowels for the AV training group between pretest and posttest and according to modality of presentation (AV, A, V).

For the A-only training group, there was a significant main effect of Time, $F(1, 19) = 63.42$, $p < .001$, $\eta_p^2 = .76$, Modality, $F(2, 38) = 23.07$, $p < .001$, $\eta_p^2 = .54$, and Vowel, $F(2, 38) = 51.90$, $p < .001$, $\eta_p^2 = .73$. The non-significant interaction of Time × Modality, $F(2, 38) = 48.51$, $p = .08$, $\eta_p^2 = .12$ indicates that, similarly to the AV group, perception in none of the modalities improved more over time than the others. Comparisons of posttest performances (see Figure 10) showed an increase of about 28% in the A modality, 30% in the AV modality and 23% in the V modality, indicating that participants improved in bimodal speech perception despite the fact that training was auditory-only. The interaction Time × Vowel was also significant $F(2, 38) = 17.72$, $p < .001$, $\eta_p^2 = .48$, with a performance increase of 45% for [ɛ̃], 20% for [ɑ̃], and 16% for [ɔ̃].

The interaction Time × Modality × Vowel was also significant, $F(4, 76) = 5.37$, $p = .001$, $\eta_p^2 = .22$. However, contrary to the AV training group, the *p* values reported in Table 11 show that for the A-only training group the gains in identification accuracy for [ɔ̃] and [ɛ̃] did not significantly differ from one another in the A and V modalities.

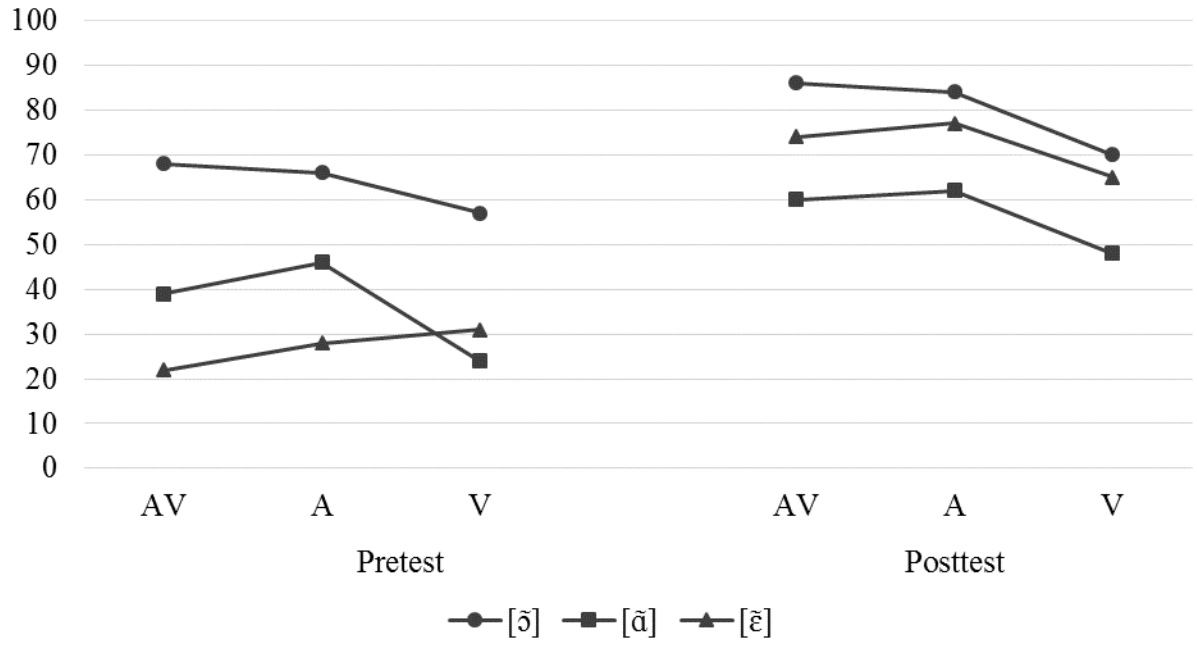

Figure 10. Changes in perceptual accuracy of French nasal vowels for the A training group between pretest and posttest and according to modality of presentation (AV, A, V).

Table 11.

Post-hoc pairwise comparisons (p values) between each vowel in the three modalities at posttest

	AV training group	A-only training group
A Modality		
[ɔ̃] - [ɑ̃]	.000	.000
[ɔ̃] - [ɛ̃]	.015	.192 (n.s)
[ɛ̃] - [ɑ̃]	.003	.008
AV Modality		
[ɔ̃] - [ɑ̃]	.001	.000
[ɔ̃] - [ɛ̃]	.019	.010
[ɛ̃] - [ɑ̃]	.004	.016
V Modality		
[ɔ̃] - [ɑ̃]	.000	.002
[ɔ̃] - [ɛ̃]	.008	.438 (n.s)
[ɛ̃] - [ɑ̃]	.003	.000

3.1.3 Comparison of training type

Accuracy scores for each of the six training sessions are displayed in Figure 11. A repeated-measures ANOVA with Time (1 to 6) as within-subjects variable and Training group as between-subjects variable was run in order to compare improvement from week to week. Results show that the Time × Training group interaction was not significant $F(5, 190) = 1.22$, $p = .301$, $\eta_p^2 = .031$. However, Time was significant $F(5, 190) = 49.35$, $p < .001$, $\eta_p^2 = .565$. Pairwise comparisons revealed that the improvement from week to week was significant ($p < .001$), except between week 5 and week 6 ($p = .079$). Thus, the data show both groups improved

similarly and consistently from session 1 to session 5, but did not make significant additional gains during session 6, the last session before the posttest.

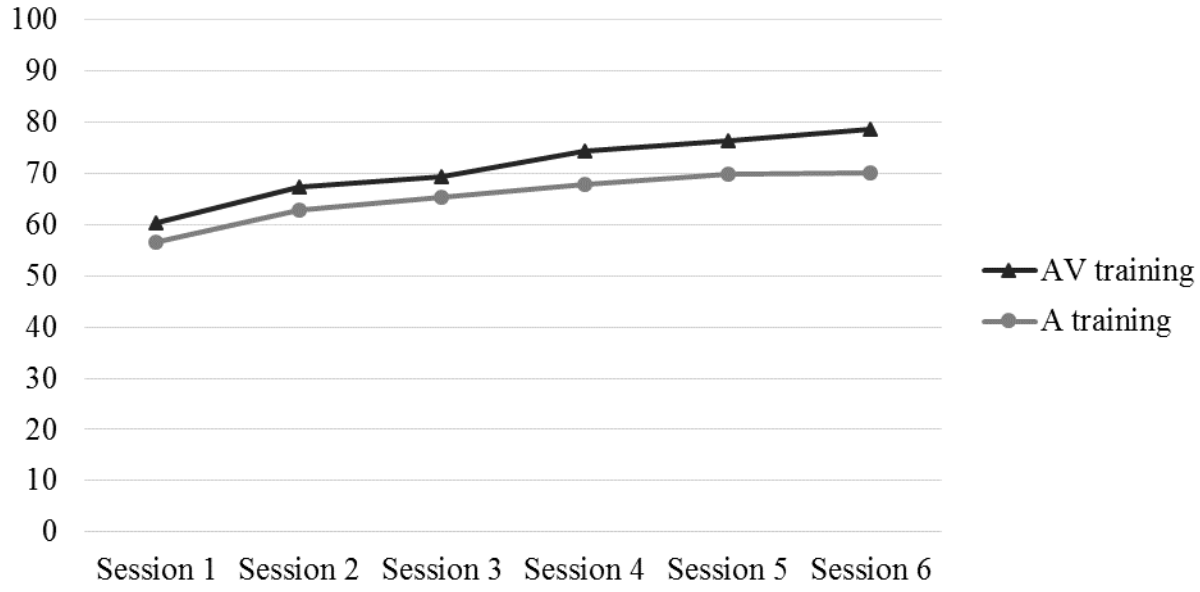

Figure 11. Percentage of correct identification per training session

A series of repeated-measures ANOVAs with Time (pretest and posttest) and Vowel (ɔ̃, ɑ̃, ɛ̃) as within-subjects factors and Training type as between-subjects factor were run to compare the effect of training condition on posttest performances for the three modalities separately. Time was a statistically significant factor in the three modalities [A modality: $F(1, 38) = 113.46$, $p < .001$, $\eta_p^2 = .740$; AV modality: $F(1, 38) = 113.14$, $p < .001$, $\eta_p^2 = .749$; V modality: $F(1, 38) = 59.14$, $p < .001$, $\eta_p^2 = .601$], but the interaction Time × Training was not statistically significant [A modality: $F(1, 38) = 0.02$, $p = .871$, $\eta_p^2 = .001$; AV modality: $F(1, 38) = 0.33$, $p = .566$, $\eta_p^2 = .009$; and V modality: $F(1, 38) = 0.05$, $p = .810$, $\eta_p^2 = .001$], indicating that neither of the training types provided greater improvement in perceptual accuracy, but both groups improved. The increase in accuracy scores for each test modality is given in Table 12.

Table 12.

Mean percentage of improvement in perceptual accuracy scores from pretest to posttest

	Test modality		
	A modality	AV modality	V modality
AV training group	29	34	25
A-only training group	28	30	23
Control group	1	5	2

Vowel was found to be a significant main effect for the three modalities [A modality: $F(2, 76) = 67.09$, $p < .001$, $\eta_p^2 = .638$; AV modality: $F(2, 76) = 91.68$, $p < .001$, $\eta_p^2 = .707$; and V modality: $F(2, 76) = 45.28$, $p < .001$, $\eta_p^2 = .544$], and the interaction Vowel × Time was also significant [A modality: $F(2, 76) = 39.43$, $p < .001$, $\eta_p^2 = .509$; AV modality: $F(2, 76) = 23.73$, $p < .001$, $\eta_p^2 = .384$; and V modality: $F(2, 76) = 8.78$, $p < .001$, $\eta_p^2 = .188$], indicating that perception of vowels improved from pretest to posttest to differing degrees (see Figures 12, 13, 14). However, no significant effects were found for the interactions Vowel × Training [A modality: $F(2, 76) = 1.32$, $p = .271$, $\eta_p^2 = .034$; AV modality: $F(2, 76) = 0.09$, $p = .913$, $\eta_p^2 = .002$; and V c modality: $F(2, 76) = 0.63$, $p = .533$, $\eta_p^2 = .016$] and Time × Vowel × Training [A modality n: $F(2, 76) = 0.11$, $p = .896$, $\eta_p^2 = .003$; AV modality: $F(2, 76) = 0.53$, $p = .590$, $\eta_p^2 = .014$; and V modality: $F(2, 76) = 0.84$, $p = .430$, $\eta_p^2 = .022$], suggesting that accurate perception varied across vowels and improved over time, but the improvement from pretest to posttest was not different between the two groups. Receiving training with audiovisual information did not lead to better perception accuracy at the posttest than receiving audio-only training.

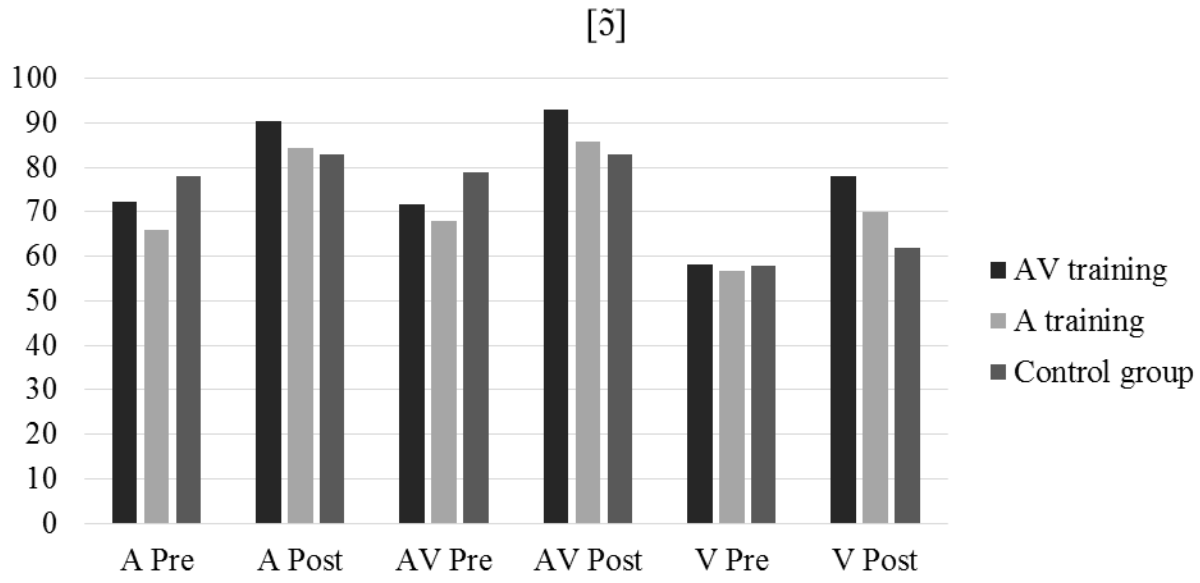

Figure 12. Mean percentage of correct identification score for [ɔ̃] according to modality of presentation (A, AV, V) and time (pretest and posttest)

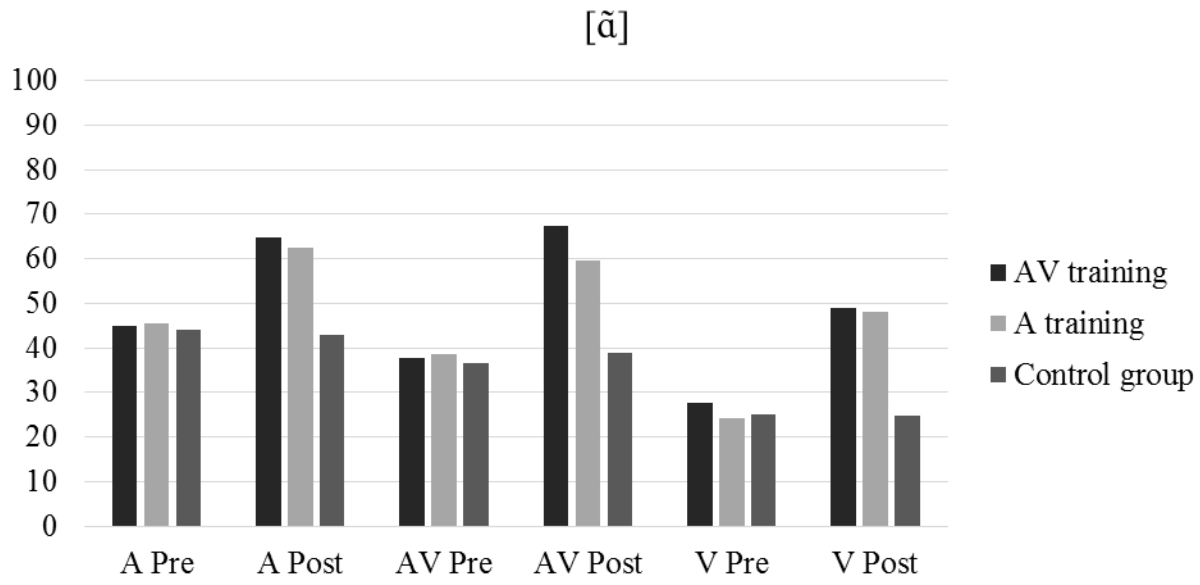

Figure 13. Mean percentage of correct identification score for [ɑ̃] according to modality of presentation (A, AV, V) and time (pretest and posttest)

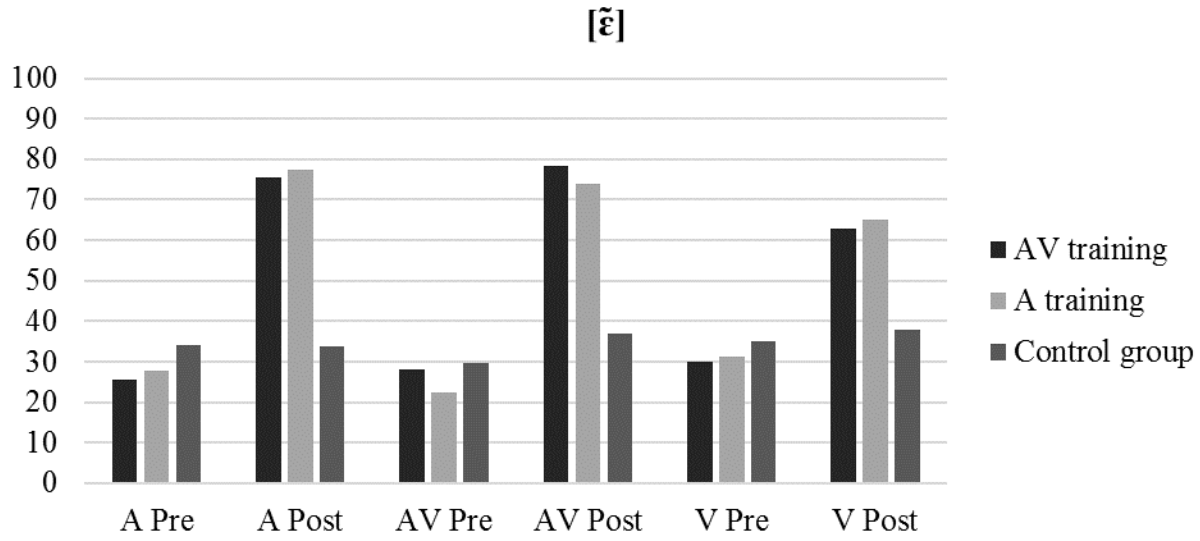

Figure 14. Mean percentage of correct identification score for [ɛ̃] according to modality of presentation (A, AV, V) and time (pretest and posttest)

The results of perception confusion among the three nasal vowels at the posttest are displayed in the confusion matrix in Table 13. The patterns of confusion for the control group are similar to those of the pretest (Table 9). For both training groups, the perception of [ɑ̃] in the A and AV modalities is now accurate more than 60% of the time, and is accurate about half the time in the V modality. Participants tended to misperceive [ɑ̃] as [ɔ̃] more frequently than as [ɛ̃], especially when only visual information was present. As noted earlier, the perception of unrounded [ɛ̃] greatly improved from pretest to posttest. This is also reflected in the pattern of confusion as trainees accurately perceived the vowel more than 74% of the time in the A and AV modalities, and almost never misperceived it as the hyper-rounded [ɔ̃].

Table 13.

Confusion matrix for vowel identification at posttest (mean percent response)

		Experimental group								
		[ɔ̃]			[ɑ̃]			[ɛ̃]		
		AV	A	C	AV	A	C	AV	A	C
	Perceived as (in %)									
A	[ɔ̃]	**90**	**85**	**85**	20	24	36	1	3	4
	[ɑ̃]	6	11	6	**65**	**62**	**44**	24	20	61
	[ɛ̃]	3	5	9	15	13	20	**76**	**77**	**35**
AV	[ɔ̃]	**93**	**86**	**84**	19	26	39	1	2	3
	[ɑ̃]	5	9	6	**67**	**60**	**40**	21	23	59
	[ɛ̃]	2	5	10	13	14	21	**78**	**74**	**38**
V	[ɔ̃]	**78**	**70**	**63**	35	40	53	3	6	3
	[ɑ̃]	12	21	14	**50**	**48**	**26**	34	30	59
	[ɛ̃]	10	9	24	16	11	21	**63**	**65**	**38**

Note. Accurate perceptions appear in bold

3.2 Research question 2: Production

To investigate whether perceptual training led to improvement in production, and whether improvement was greater with AV perceptual training than A training, recordings were made of the participants' productions at pretest and posttest. Stimuli consisted of the 108 tokens from the perception pretest and posttest and were judged by two native French raters. The results are organized in the following manner: first, I will report the results of a forced identification

task where raters were asked to identify the vowel produced by the L2 speakers by choosing one of the three nasal vowels. Production tokens which were accurately identified by the native raters were coded as correct, whereas tokens which were inaccurately identified were coded as incorrect, regardless of which other vowel the raters chose. Then, I will report the results of a quality rating task where raters judged the quality of the L2 speakers' production of the vowel on a scale from 1 (terrible) to 7 (excellent).

3.2.1 Identification rating

The raters' pretest and posttest scores were totaled for each participant and mean ratings for each experimental group are displayed in Figure 15. A one-way ANOVA with Group (AV training, A-only training, Control) as between-subjects variable and Pretest scores as independent variable revealed that the three groups were not equal at the pretest, $F(2, 6040) = 7.20$, $p = .001$, $\eta^2 = .002$. LSD post-hoc tests indicated that the score of the control group was higher than the AV training group ($p < .001$) and the A-only training group ($p = .033$), but that the two training groups were comparable at pretest ($p = .08$).

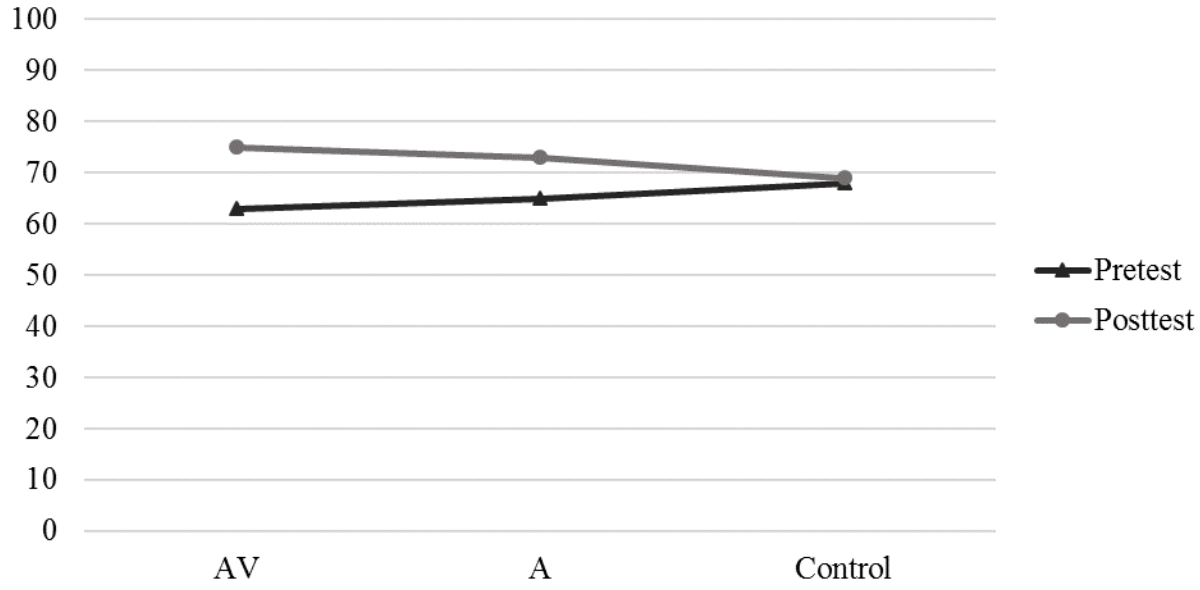

Figure 15. Percentage of accurate production at the pretest and posttest as rated by native French speakers

A repeated-measures ANOVA was used for each group separately to test whether training affected accuracy in pronouncing French nasal vowels. Results show that the productions of the two training groups significantly improved from the pretest to the posttest, [$F_{AV}(1, 1833) = 104.12$, $p < .001$, $\eta_p^2 = .054$; $F_A(1, 2110) = 40.81$, $p < .001$, $\eta_p^2 = .019$], but that the control group did not improve, $F(1, 2046) = 1.30$, $p = .254$, $\eta_p^2 = .001$, indicating that perceptual training leads to improvement in production. Change scores across groups were computed in order to further tell whether one type of training was better than the other in improving pronunciation. A one-way ANOVA with Group as between-subjects factor and Score change as independent variable showed that there was a significant effect of Group, $F(2, 5991) = 24.16$, $p < .001$, and a Scheffé post-hoc test revealed significant differences across the three groups (Figure 16). The change in improvement was greater for the AV group than for the A group ($p = .002$) and for the control group ($p < .001$), and the A group improved more than the control group, which did not show significant improvement ($p = .002$).

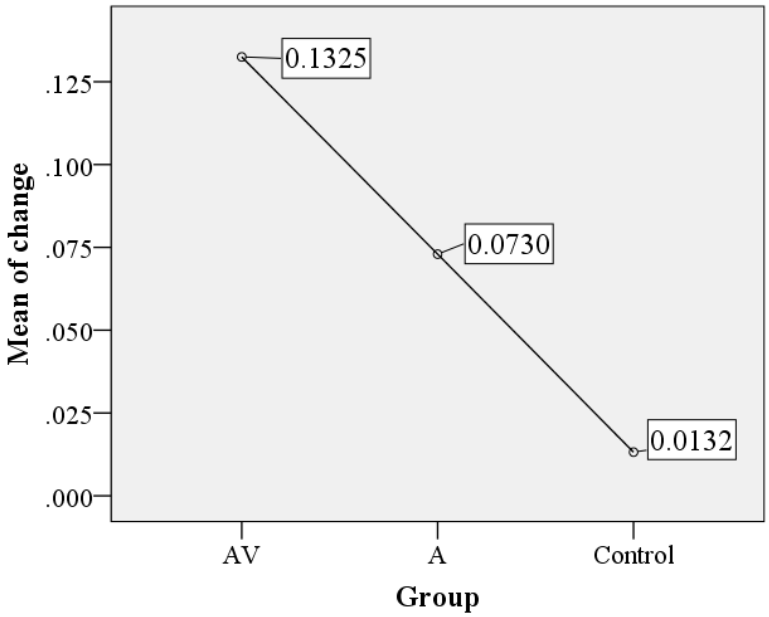

Figure 16. Mean of change from production pretest to posttest according to identification rating of native perceivers

In order to see whether the production of some vowels improved more than others did and whether improvement varied across training modalities, a series of one-way ANOVAs followed by Scheffé post-hoc tests were run for each vowel separately, with Group as between-subjects factor and Score change as dependent variable. Results revealed that there was a significant effect of group for [ɔ̃], $F(2, 1996) = 15.63$, $p < .001$ (Figure 17). The AV training group improved significantly more than the A-only training group ($p = .014$) and the control group ($p < .001$) did, and the A-only training group's change in production was significantly greater than the control group's change ($p = .021$).

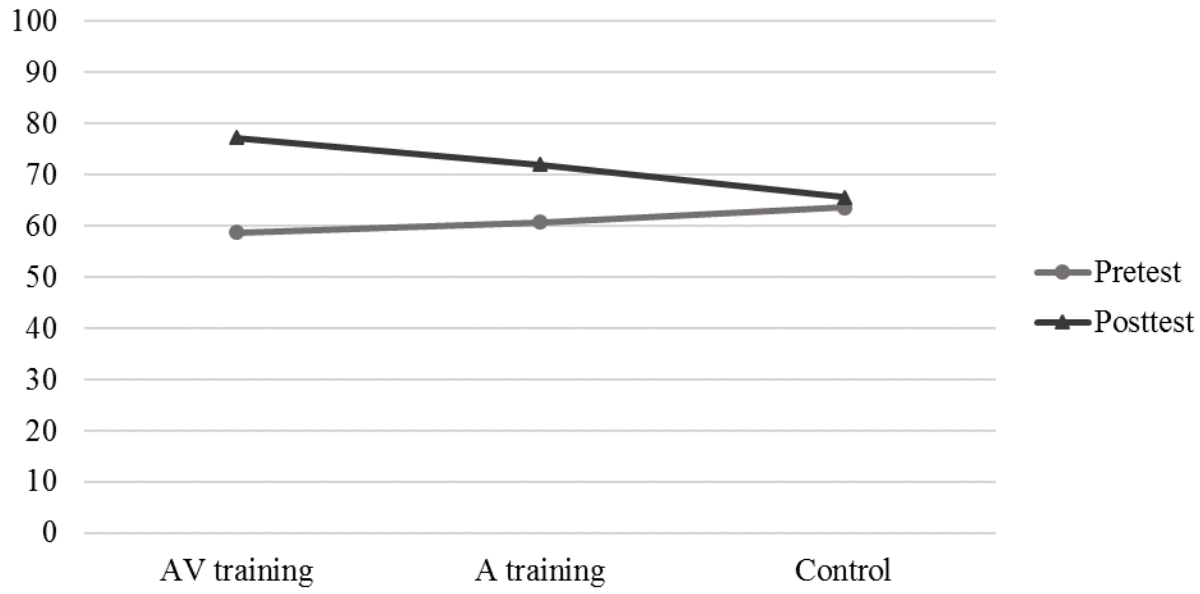

Figure 17. Percentage of accurate production for [ɔ̃] at the pretest and posttest

Results for [ɑ̃] showed that there was again a significant effect of Group, $F(2, 1997) = 3.69$, $p = .025$ (Figure 18). The difference in improvement change was not statistically different between the AV and A-only training groups ($p = .59$) or between the A-only training and control groups ($p = .225$), but the AV training group's change in production was significantly greater than the control group's change ($p = .028$). Note that the production performance of the control

group actually decreased from pretest to posttest, although the difference was not statistically significant (p = .08).

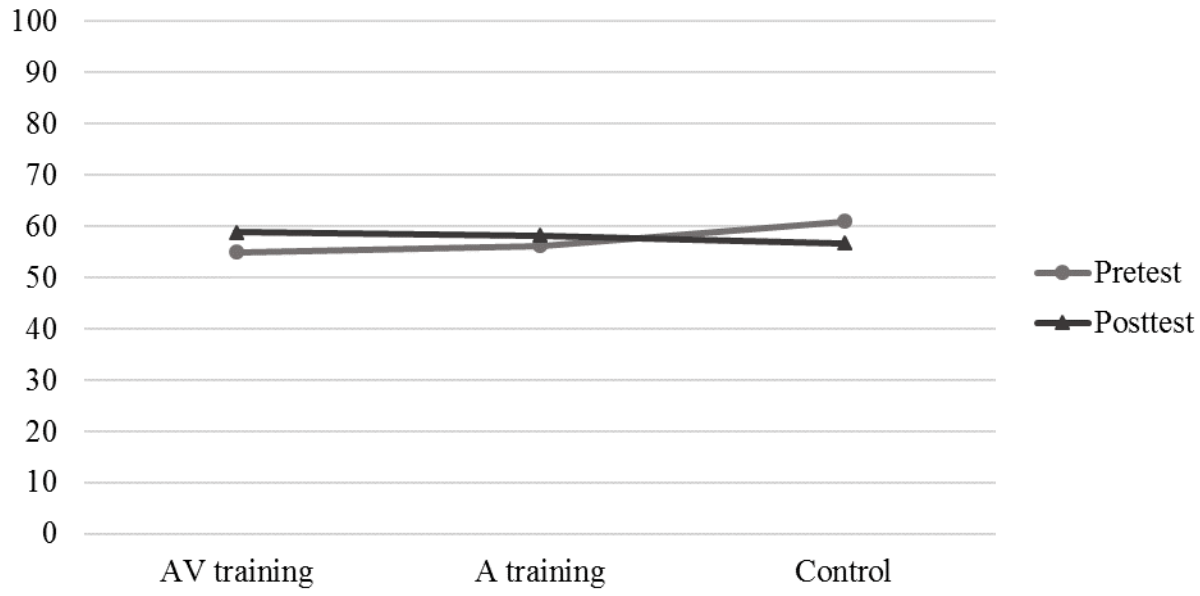

Figure 18. Percentage of accurate production for [ɑ̃] at the pretest and posttest

Results for [ɛ̃] are displayed in Figure 19. Again, there was a significant effect of Group, $F(2, 1996) = 8.99$, $p < .001$, and post-hoc tests showed that the AV training group improved significantly more than the A-only training (p = .046) and the control group ($p < .001$) did. The change in improvement was not statistically significant between the A-only training and the control group (p = .185).

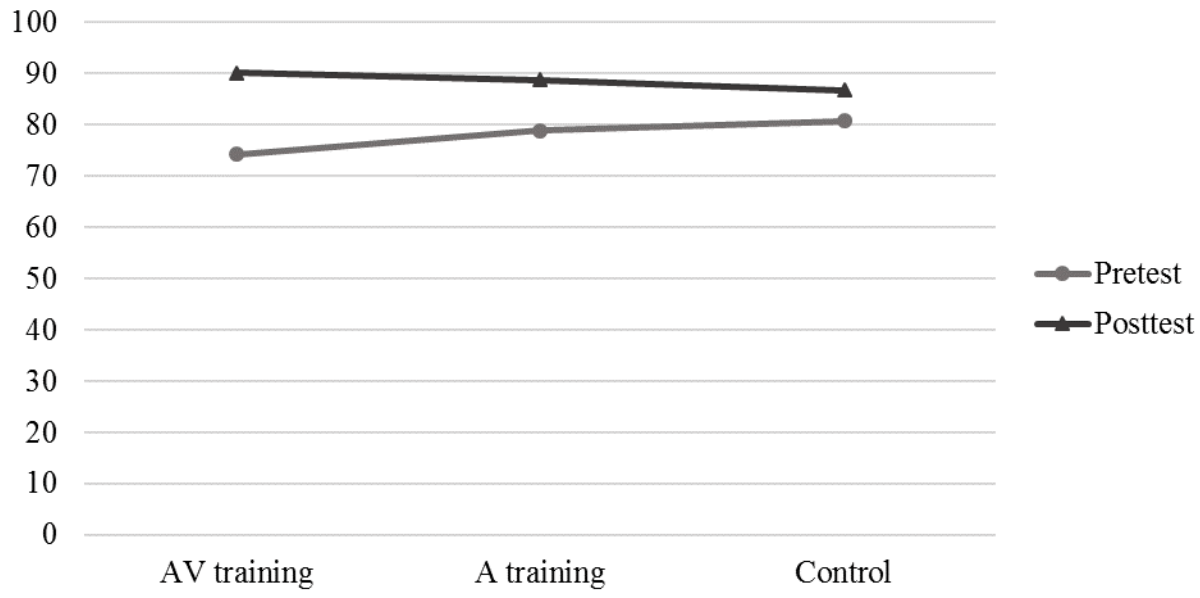

Figure 19. Percentage of accurate production for [ɛ̃] at the pretest and posttest

3.2.2 Quality rating

The pretest and posttest scores of the raters were totaled for each participant and mean ratings for each experimental group are displayed in Table 14. A one-way ANOVA with Group (AV training, A-only training, Control) as between-subjects variable and Pretest scores as independent variable revealed that, similarly to the identification rating, the three groups were not equal at the pretest, $F(2, 6040) = 14.53$, $p < .001$. LSD post-hoc tests indicated that the score of the control group was higher than the ones of AV training group ($p < .001$) and the A-only training group ($p < .001$), but that the two training groups were equal at pretest ($p = .41$).

Table 14.

Mean production ratings in pretest and posttest per group (7-point scale)

	AV training group	A-only training group	Control group
Pretest score	4.91 (1.62)	4.95 (1.61)	5.16 (1.56)
Posttest score	5.41 (1.60)	5.23 (1.75)	5.44 (2.06)

Note. Standard deviations are in parentheses

A repeated-measures ANOVA was used for each group separately to test whether training affected the quality of pronunciation of the French nasal vowels. Results show that the productions of the two training groups significantly improved from the pretest to the posttest, [$F_{AV}(1, 1833) = 152.73$, $p < .001$, $\eta_p^2 = .077$; $F_A(1, 2157) = 43.00$, $p < .001$, $\eta_p^2 = .020$], and that the overall quality of pronunciation of the control group improved as well, $F(1, 2048) = 33.91$, $p < .001$, $\eta_p^2 = .016$. In order to further tell whether one group was better than the others in improving their pronunciation, change scores across groups were computed (Figure 20). A one-way ANOVA with Group as between-subjects factor and Score change as independent variable showed that there was a significant effect of group, $F(2, 6040) = 7.70$, $p < .001$. A Scheffé post-hoc test further revealed that the change in improvement was greater for the AV group than for the A group ($p = .003$) and the control group ($p = .002$), and the A group did not improve more than the control group did ($p = .995$).

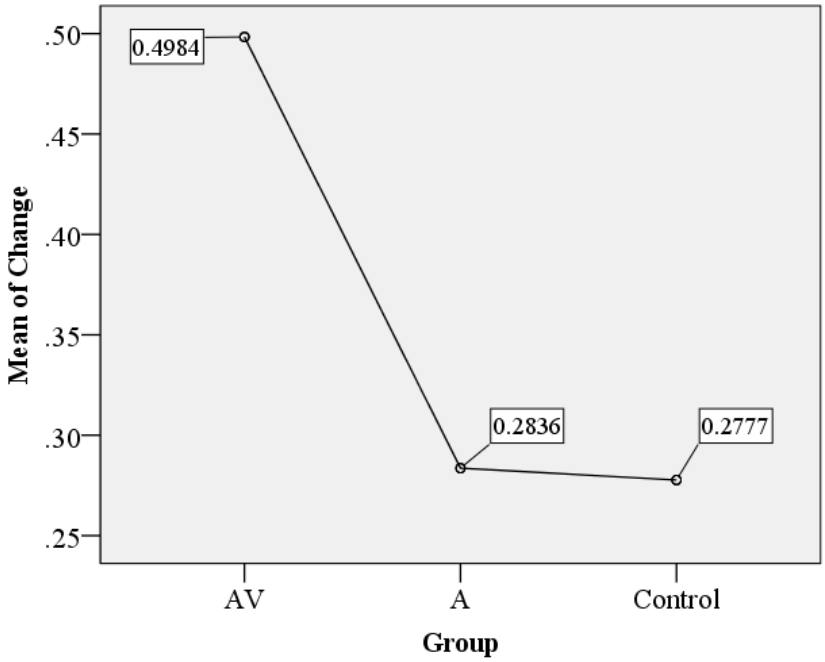

Figure 20. Mean change from production pretest to posttest according to the quality rating by native perceivers

In order to see whether the quality of production of some vowels improved more than others and whether improvement varied across modalities, a series of one-way ANOVAs followed by Scheffé post-hoc tests were run for each vowel separately, with Group as between-subjects factor and Score change as independent variable. Results revealed that there was a significant effect of group for [ɔ̃], $F(2, 2010) = 11.15$, $p < .001$ (Figure 21). The improvement for the two training groups did not significantly differ from each other ($p = .094$), but both training groups had significantly greater production improvement than the control group (AV: $p < .001$, A: $p = .029$).

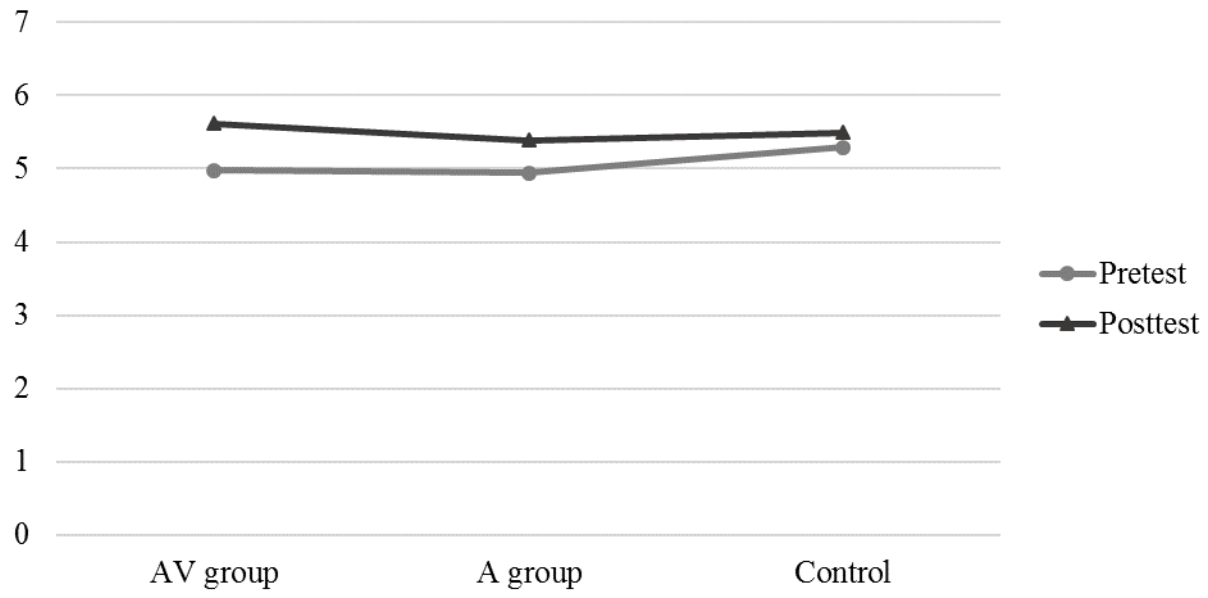

Figure 21. Mean of production rating (7-point scale) for [õ] at the pretest and posttest

Results for [ɑ̃] are displayed in Figure 22. Again, there was a significant effect of Group, $F(2, 2013) = 3.83$, $p = .022$, and post-hoc tests showed that the AV training group improved significantly more than the A-only training group ($p = .022$) whose performance actually decreased from 4.82 at the pretest to 4.74 at the posttest. Nevertheless, the change in improvement between the control group and the two training groups was not statistically significant (AV: $p = .352$, A: $p = .405$).

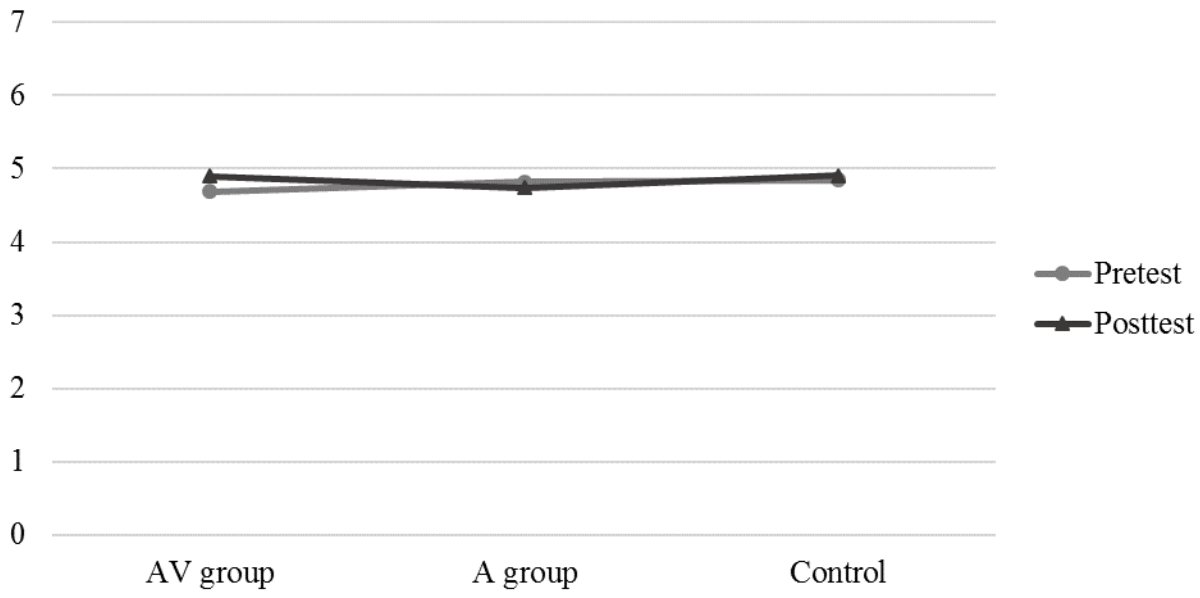

Figure 22. Mean of production rating (7-point scale) for [ɑ̃] at the pretest and posttest

Results for [ɛ̃] showed that there was no significant effect of group, $F(2, 2013) = 0.70$, $p = .494$ (Figure 23), indicating that none of the groups improved their pronunciation of the vowel [ɛ̃] more than the other groups, as measured by a quality rating task.

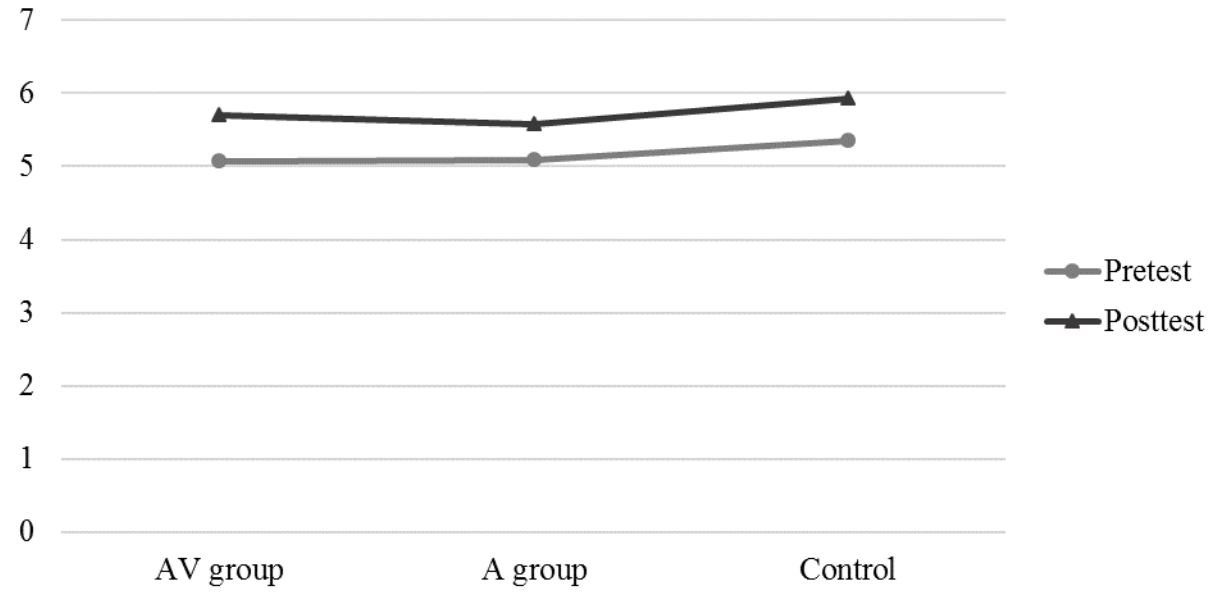

Figure 23. Mean of production rating (7-point scale) for [ɛ̃] at the pretest and posttest

3.3 Research question 3: Consonantal context

A possible factor that could affect the perception of the vowels is the consonantal context. The third research question addressed in this dissertation therefore aimed at analyzing whether the preceding and following consonants had an effect on the perception of the three nasal vowels, and whether this effect was the same for the two training groups (AV and A-only). Recall that in the perception posttest, the consonants used were [p-t-k-b-d-g-s-z-f-v-ʃ-ʒ].

This section is organized the following way: to begin with, I analyzed the effects of the initial consonant before turning to the data for the final consonant. First, I compared the effect of labiality of the initial consonant in the three different modalities of testing (AV, A, V) for each of the nasal vowels. The bilabial [p-b], the labiodental [f-v], and the palatal [ʃ-ʒ] are labial consonants (i.e., for the first two, the lips are an active articulator and protrusion of the lips is necessary for the latter), whereas the dental [t-d], the velar [k-g], and the alveolar [s-z] are non-labial consonants. Then, I further analyzed the effect of place of articulation in each modality (AV, A, V) and for each nasal vowel by comparing the six places of articulation mentioned earlier. Finally, the two manners of articulation (occlusive vs. fricative) are compared in each of the modalities and for each vowel.

3.3.1 The effect of labiality of the initial consonant

The main accuracy scores according to the labiality of the initial consonant are displayed in Table 15. A series of repeated-measures ANOVAs with Labiality (labial and non-labial) and Vowels as within-subjects factor, Training type as between-subjects factor, and Posttest score as dependent variable were run to compare the effect of consonantal context in the three testing modalities (AV, A, V) separately.

Table 15.

Mean percentage of accuracy score at the perception posttest according to labiality, testing modality, and training group

Vowel	Labiality	AV modality		A modality		V modality	
		AV training	A training	AV training	A training	AV training	A training
ɔ̃	Labial	91.1 (15.1)	83.1 (17.7)	89.7 (11.7)	80.0 (18.7)	77.8 (21.4)	69.4 (27.4)
	Non labial	94.7 (8.3)	88.6 (9.6)	90.8 (9.4)	88.9 (12.3)	78.6 (18.1)	69.7 (22.1)
ɑ̃	Labial	65.3 (29.9)	58.6 (30.8)	61.9 (27.4)	63.6 (30.7)	45.0 (22.3)	47.8 (20.8)
	Non labial	69.2 (30.8)	60.6 (33.3)	67.5 (27.3)	61.1 (26.7)	53.9 (25.3)	48.3 (26.9)
ɛ̃	Labial	78.3 (24.8)	74.2 (22.5)	73.6 (28.4)	78.9 (25.6)	60.6 (25.9)	61.1 (24.2)
	Non labial	78.3 (24.7)	73.9 (23.6)	77.2 (25.6)	75.8 (22.8)	64.4 (27.1)	68.3 (25.1)

Note. Standard deviations are in parentheses

Results show that Vowel had a significant main effect for the three modalities [A modality: $F(2, 76) = 23.05$, $p < .001$, $\eta_p^2 = .378$; AV modality: $F(2, 76) = 24.74$, $p < .001$, $\eta_p^2 = .394$; and V c modality: $F(2, 76) = 24.16$, $p < .001$, $\eta_p^2 = .389$], but that the effect of Training Group [A modality: $F(1, 38) = 0.13$, $p = .720$, $\eta_p^2 = .003$; AV modality: $F(1, 38) = 1.21$, $p = .278$, $\eta_p^2 = .031$; and V modality: $F(1, 38) = 0.20$, $p = .657$, $\eta_p^2 = .005$] and the interaction Vowel × Training group [A modality: $F(2, 76) = 0.61$, $p = .543$, $\eta_p^2 = .016$; AV modality: $F(2, 76) = 0.11$, $p = .890$, $\eta_p^2 = .003$; and V modality: $F(2, 76) = 1.15$, $p = .322$, $\eta_p^2 = .029$] were not

significant, indicating that perception accuracy differed across vowels, but that both groups performed similarly (see Figures 24 to 26).

Labiality was found to have a significant effect in the A modality [$F(1, 38) = 4.27$, $p = .045$, $\eta_p^2 = .101$] and V modality [$F(1, 38) = 7.24$, $p = .011$, $\eta_p^2 = .160$] and approached significance in the AV modality [$F(1, 38) = 3.76$, $p = .060$, $\eta_p^2 = .090$], but the interaction Labiality × Training group was not significant [A modality: $F(1, 38) = 1.11$, $p = .298$, $\eta_p^2 = .028$; AV modality: $F(1, 38) = 0.01$, $p = .971$, $\eta_p^2 = .000$; and V modality: $F(1, 38) = 0.47$, $p = .494$, $\eta_p^2 = .012$]. In addition, the interaction Labiality × Vowel was also not significant [A modality: $F(2, 76) = 1.10$, $p = .336$, $\eta_p^2 = .028$; AV modality: $F(2, 76) = 0.84$, $p = .435$, $\eta_p^2 = .022$; and V modality: $F(2, 76) = 0.97$, $p = .384$, $\eta_p^2 = .025$]. Therefore, the data show that perception was better with non-labial consonants across all modalities and vowels, and that, this effect held true regardless of one's training condition.

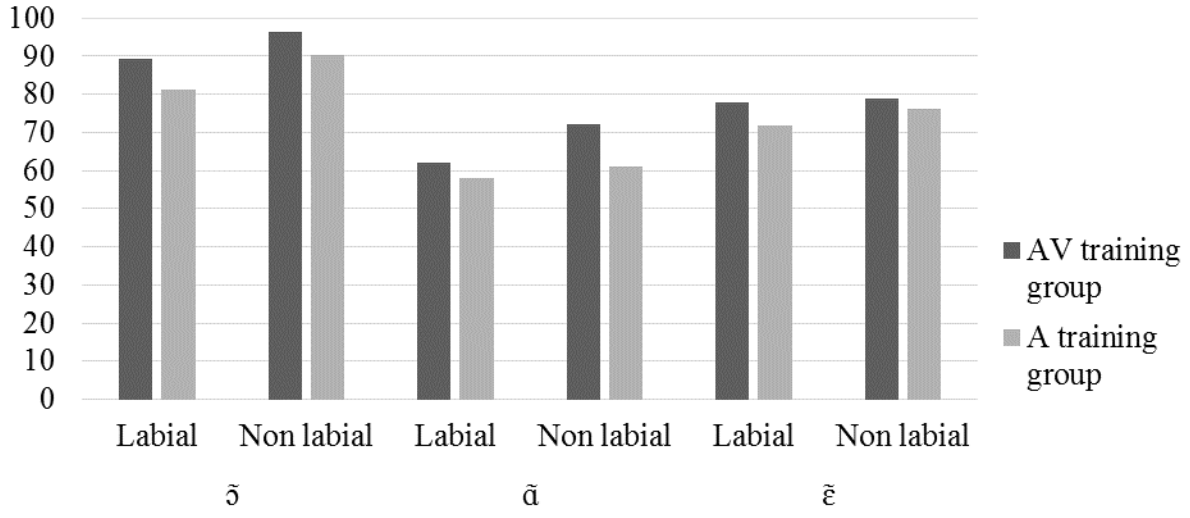

Figure 24. Percentage of accuracy score at the AV perception posttest according to labiality

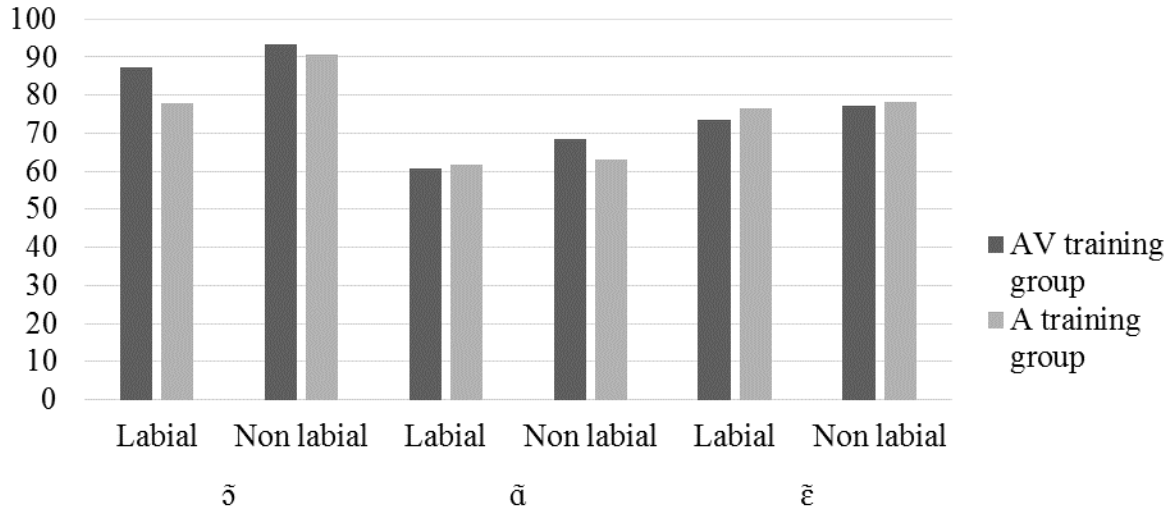

Figure 25. Percentage of accuracy score at the A-only perception posttest according to labiality

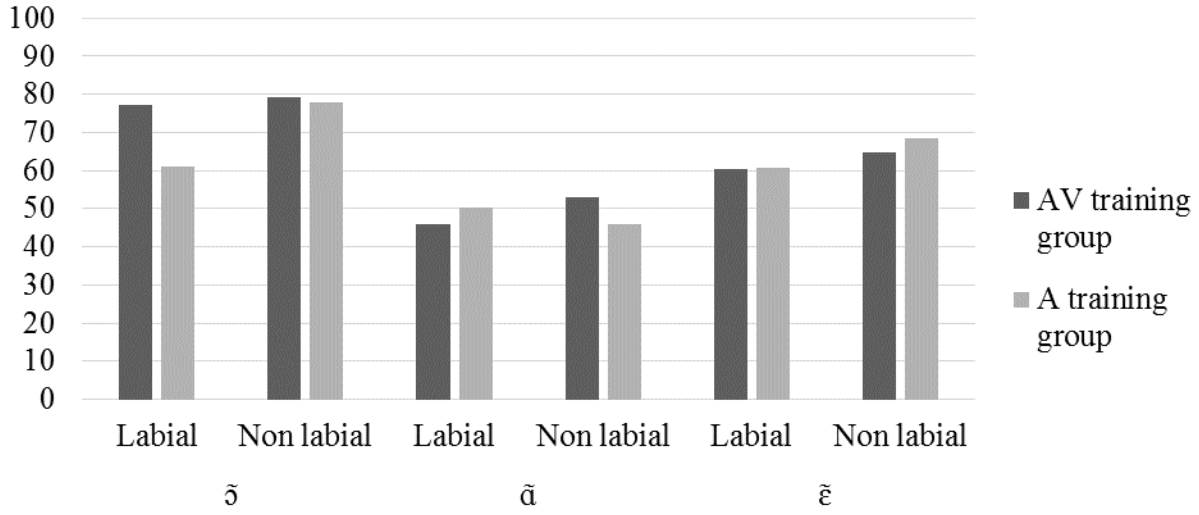

Figure 26. Percentage of accuracy score at the V-only perception posttest according to labiality

3.3.2 The effect of place of articulation of the initial consonant

The percentages of accuracy scores according to place of articulation, modality, and training group are shown in Table 16. To further investigate the effect of consonantal context, and more particularly the effect of the place of articulation of the initial consonant, a series of repeated-measures ANOVAs were conducted for each modality (AV, A, V). The between-

subjects variable was Training group (A, AV) and the within-subjects variables were Vowels (ɔ̃, ɑ̃, ɛ̃) and Place of articulation (bilabial, dental, velar, labiodental, alveolar, palatal).

Table 16.

Mean percentage of accuracy score at the perception posttest according to place of articulation, test modality, and training group

Vowel	Labiality	AV modality		A modality		V modality	
		AV training	A training	AV training	A training	AV training	A training
ɔ̃	Bilabial	97.5 (6.0)	93.3 (11.3)	95.8 (11.8)	90.8 (15.7)	83.3 (25.8)	89.2 (14.5)
	Dental	95 (12.2)	89.2 (13.5)	94.2 (8.0)	90 (14.7)	80 (23.8)	75 (23.8)
	Velar	96.7 (6.8)	88.3 (13.3)	90 (16.5)	91.7 (11.3)	74.2 (23.8)	70 (21.3)
	Labiodental	91.7 (16.7)	86.7 (17.7)	91.7 (12.7)	87.5 (22.8)	80 (20.5)	60 (38.7)
	Alveolar	92.5 (14.7)	88.3 (13.3)	88.3 (16.2)	85 (18.5)	81.7 (20.8)	64.2 (33.3)
	Palatal	84.2 (26.7)	68.3 (31.0)	81.7 (21.5)	61.7 (28.0)	70 (28.3)	59.2 (38.3)
ɑ̃	Bilabial	70 (30.3)	60 (33.8)	65.8 (29.3)	64.2 (32.0)	45 (27.0)	43.3 (24.3)
	Dental	71.7 (34.5)	61.7 (35.0)	70 (28.8)	60 (33.5)	49.2 (29.8)	45 (32.3)
	Velar	75 (29.3)	61.7 (32.8)	70 (30.8)	65 (29.5)	65 (31.0)	49.2 (28.3)
	Labiodental	60 (34.7)	54.2 (37.0)	55.8 (34.2)	63.3 (39.5)	45.8 (23.3)	49.2 (25.5)
	Alveolar	60.8 (34.7)	58.3 (37.5)	62.5 (31.0)	58.3 (27.2)	47.5 (28.2)	50.8 (31.2)
	Palatal	65.8 (32.5)	61.7 (32.3)	64.2 (31.2)	63.3 (30.3)	44.2 (27.2)	50.8 (28.3)

Table 16 (cont'd)

ɛ̃	Bilabial	79.2 (25.8)	76.7 (21.8)	72.5 (32.5)	80.8 (27.2)	65.8 (33.0)	66.7 (32.8)
	Dental	76.7 (32.2)	66.7 (29.5)	78.3 (31.0)	71.7 (25.3)	64.2 (33.3)	67.5 (30.7)
	Velar	80.8 (31.5)	85 (25.2)	80.8 (31.5)	82.5 (21.8)	64.2 (32.5)	71.7 (28.5)
	Labiodental	78.3 (25.3)	70.8 (28.5)	72.5 (31.5)	77.5 (27.2)	56.7 (29.7)	61.7 (31.0)
	Alveolar	77.5 (24.8)	70 (32.7)	71.7 (30.2)	73.3 (30.2)	65 (24.0)	65 (25.5)
	Palatal	77.5 (29.7)	75 (23.8)	75.8 (27.2)	78.3 (30.0)	59.2 (33.5)	55 (32.8)

Note. Standard deviations are in parentheses

Results for the AV test modality (Figure 27) show that there were significant main effects for Vowel [$F(2, 76) = 24.74$, $p < .001$, $\eta_p^2 = .394$], and Place of articulation [$F(5, 190) = 6.64$, $p < .001$, $\eta_p^2 = .149$], and that the interaction Vowel × Place of articulation was significant [$F(10, 380) = 2.92$, $p = .002$, $\eta_p^2 = .072$]. However, there was no significant effect for Training group [$F(1, 38) = 1.21$, $p = .278$, $\eta_p^2 = .031$], and the interactions Vowel × Training group [$F(2, 76) = 0.11$, $p = .89$, $\eta_p^2 = .003$], Place of articulation × Training group [$F(5, 190) = 0.27$, $p = .92$, $\eta_p^2 = .007$], and Place of articulation × Vowel × Training group [$F(10, 380) = 1.01$, $p = .43$, $\eta_p^2 = .026$] were also not significant. This indicates that the performances of the two training groups were similar across vowels and places of articulations, and that therefore training that incorporated visual cues did not lead to better performance than training with audio information only.

We already know from the results of the first research question that higher perception scores were obtained for [ɔ̃], followed by [ɑ̃] and finally [ɛ̃] across the three testing modalities. Averaging across vowels, the scores were higher for velar (81.2%), followed by bilabial (79.4%), dental (76.8%), alveolar (74.6%), labiodental (73.7%), and finally palatal (72.1%). Regarding

places of articulation in the AV modality, pairwise comparisons found significant differences between velar and alveolar (p = .059), velar and labiodental (p = .019) and velar and palatal (p = .003), and between bilabial and labiodental (p = .010), bilabial and palatal (p = .004), and a difference approaching significance between bilabial and alveolar (p = .078). Follow-up analyses revealed that place of articulation did not affect perception for the AV group, but led to significant differences for the A group only with the vowel [ɔ̃]. As displayed in Figure 27, the perception accuracy for vowels in the contexts following palatal consonants was significantly lower than the one for bilabial (p = .005), labiodental (p = .041) and alveolar (p = .038).

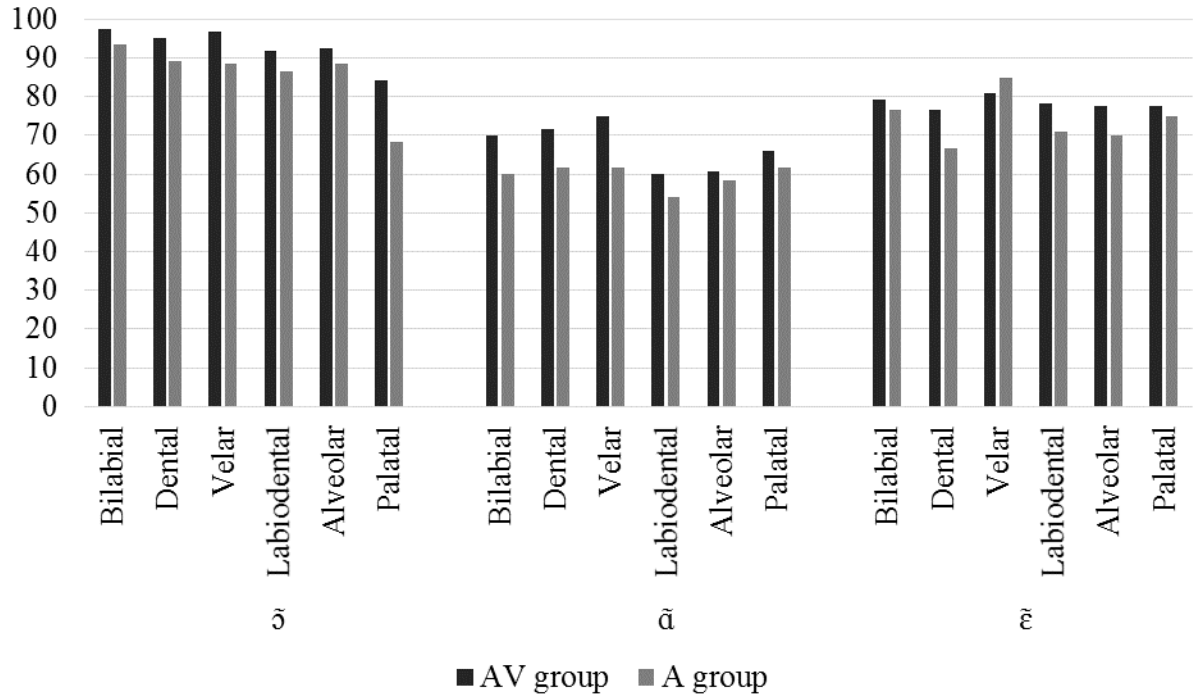

Figure 27. Percentage of accuracy score in the AV test modality at the perception posttest according to the place of articulation and vowel

Results for the A-only test modality (Figure 28) show that there were significant main effects for Vowel [$F(2, 76) = 23.05$, $p < .001$, $\eta_p^2 = .378$], and Place of articulation [$F(5, 190) = 6.53$, $p < .001$, $\eta_p^2 = .147$], and that the interaction Vowel × Place of articulation was significant [$F(10, 380) = 3.26$, $p < .001$, $\eta_p^2 = .079$]. However, there was no significant effect for Training

group [$F(1, 38) = 0.13$, $p = .720$, $\eta_p^2 = .003$], and the interactions Vowel × Training group [$F(2, 76) = 0.61$, $p = .54$, $\eta_p^2 = .016$] and Place of articulation × Vowel × Training group [$F(10, 380) = 1.19$, $p = .29$, $\eta_p^2 = .030$] were also not significant. The interaction Place of articulation × Training group approached significance [$F(5, 190) = 2.03$, $p = .075$, $\eta_p^2 = .051$]. This indicates that, type of vowel and place of articulation had a similar effect on the two training groups.

Averaging across vowels, the scores were higher for velar (80%), followed by bilabial (78.3%), dental (77.3%), labiodental (74.7%), alveolar (73.3%), and finally palatal (70.3%). Pairwise comparisons found significant differences between velar and alveolar ($p = .045$), and velar and palatal ($p = .001$), and between bilabial and palatal ($p = .002$) and dental and palatal ($p = .010$). Follow up analyses revealed that the interaction Place of articulation × Training group was only significant for [ð̃] ($p = .042$). Place of articulation did not affect perception for the AV group, but led to significant differences for the A group. As displayed in Figure 28, the perception accuracy for vowels in contexts following palatal consonants was significantly lower than all the other consonants (at level $p = .001$).

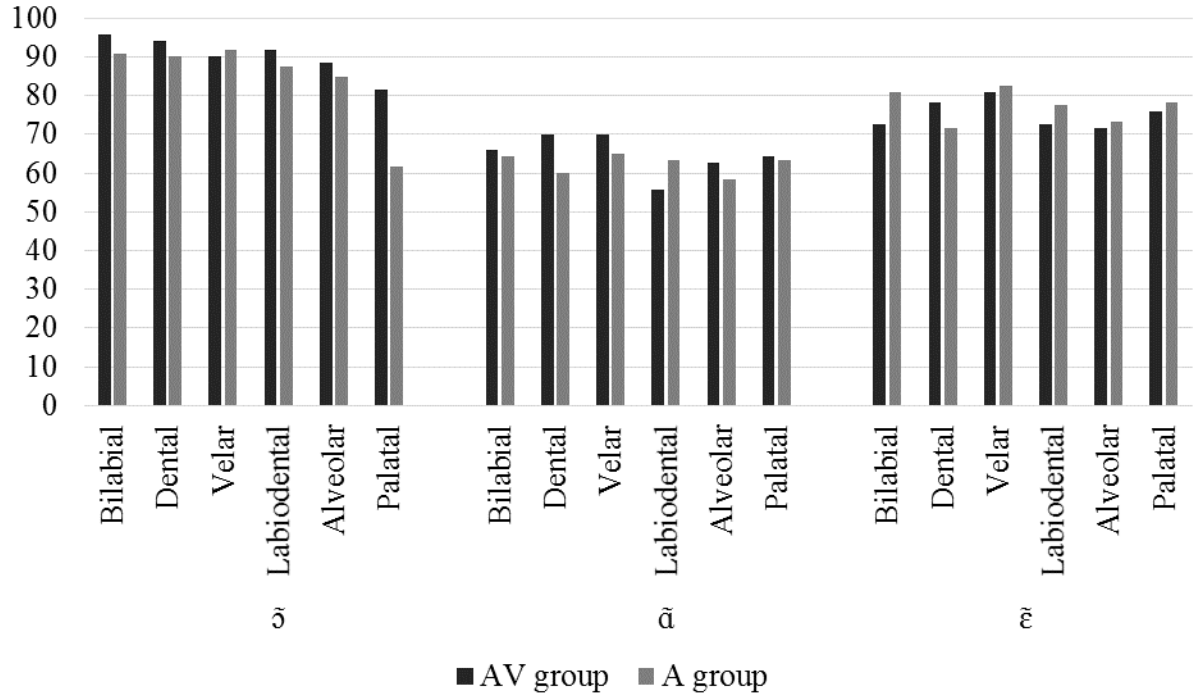

Figure 28. Percentage of accuracy score in the A-only test modality at the perception posttest according to the place of articulation and vowel

Results for the V-only test modality (Figure 29) show that there were significant main effects for Vowel [$F(2, 76) = 24.16$, $p < .001$, $\eta_p^2 = .389$], and Place of articulation [$F(5, 190) = 4.51$, $p = .001$, $\eta_p^2 = .106$], and that the interactions Vowel × Place of articulation [$F(10, 380) = 2.65$, $p = .004$, $\eta_p^2 = .065$] and Place of articulation × Vowel × Training group [$F(10, 268.8) = 2.04$, $p = .049$, $\eta_p^2 = .051$] were significant. However, there was no significant main effect for Training group [$F(1, 38) = 0.20$, $p = .65$, $\eta_p^2 = .005$], and the interactions Vowel × Training group [$F(2, 76) = 1.15$, $p = .32$, $\eta_p^2 = .029$] and Place of articulation × Training group [$F(5, 190) = 0.04$, $p = .83$, $\eta_p^2 = .011$] were also not significant. This indicates that the initial consonant had different effects for certain vowels, and that these effects were different between the two groups.

Averaging across vowels, the scores were higher for velar (65.7%), followed by bilabial (65.5%), dental (63.5%), alveolar (62.5%), labiodental (58.9%), and finally palatal (56.4%). Pairwise comparisons only found a significant difference between velar and palatal ($p = .041$).

Follow up analyses revealed that the interaction Place of articulation × Training group was only significant for [ɔ̃] (p = .023). Place of articulation did not affect perception for the AV group, but led to significant differences for the A group. As displayed in Figure 29, the perception accuracy for vowels in the contexts following bilabial consonants was significantly higher than all the other consonants. The difference between the dental and palatal was also significant (p = .025).

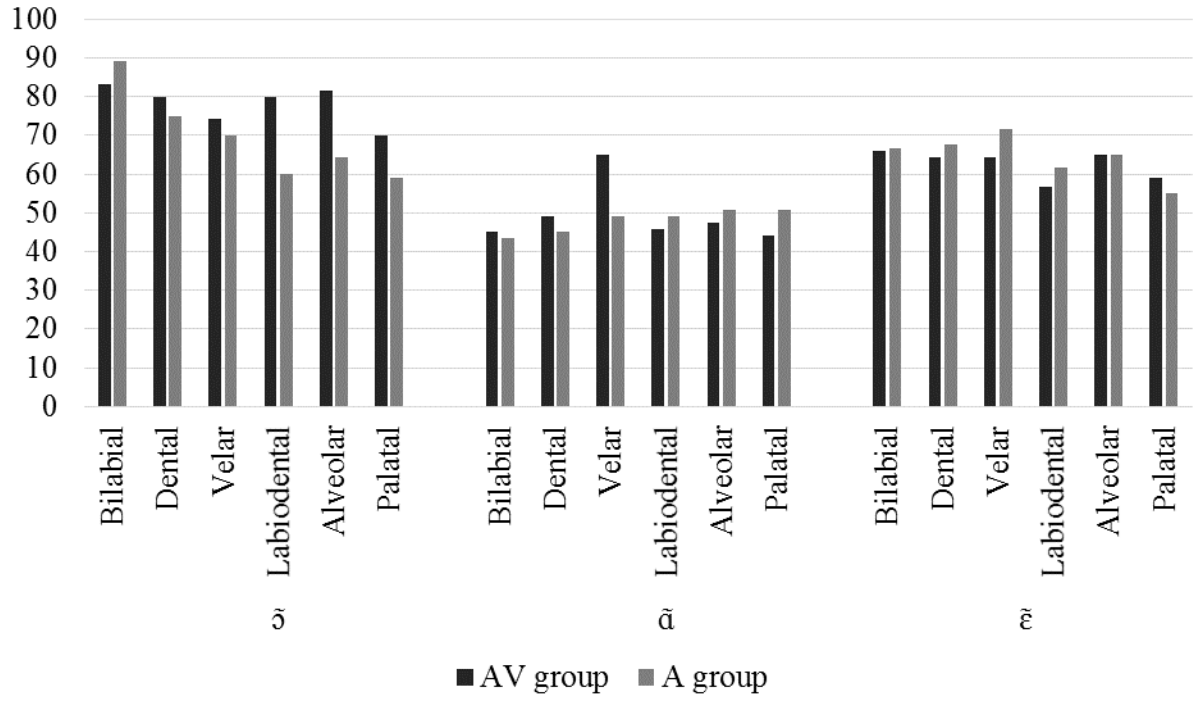

Figure 29. Percentage of accuracy score in the V-only test modality at the perception posttest according to the place of articulation and vowel

3.3.3 The effect of the initial consonant's manner of articulation

The main accuracy scores displayed in Table 17 show that accurate perception in the three testing modalities was higher with occlusive initial consonants, except in the V modality for the perception of [ɑ̃] by the A-only training group. A series of repeated-measures ANOVAs with Manner of articulation as a within-subjects factor, Training type as a between-subjects factor, and Posttest score as dependent variable were run to compare the effect of the manner of articulation in the three test modalities separately.

Table 17.

Mean percentage of accuracy score at the perception posttest according to manner of articulation, testing modality, and training group

Vowel	Manner	AV modality		A-only modality		V modality	
		AV training	A training	AV training	A training	AV training	A training
ɔ̃	Fricative	89.4 (18.3)	81.4 (17.7)	87.2 (13.1)	78.1 (19.6)	77.2 (20.4)	61.1 (34.5)
	Occlusive	96.4 (5.7)	90.3 (10.4)	93.3 (8.6)	90.8 (11.4)	79.2 (18.9)	78.1 (16.9)
ɑ̃	Fricative	62.2 (31.6)	58.1 (30.6)	60.8 (28.3)	61.7 (28.9)	45.8 (23.4)	50.3 (24.3)
	Occlusive	72.2 (29.0)	61.1 (31.2)	68.6 (26.5)	63.1 (26.7)	53.1 (25.7)	45.8 (23.9)
ɛ̃	Fricative	77.8 (24.9)	71.9 (23.6)	73.6 (27.3)	76.4 (27.1)	60.3 (22.3)	60.8 (23.0)
	Occlusive	78.9 (24.9)	76.1 (19.1)	77.2 (26.1)	78.3 (20.8)	64.7 (29.4)	68.6 (25.6)

Note. Standard deviations are in parentheses

Results show that, in the AV modality (Figure 30), there was a significant main effect of Manner for [ɔ̃], $F(1, 38) = 12.20$, $p = .001$, $\eta_p^2 = .243$, and [ɑ̃], $F(1, 38) = 9.53$, $p = .004$, $\eta_p^2 = .201$, but not for [ɛ̃], $F(1, 38) = 2.69$, $p = .109$, $\eta_p^2 = .066$, indicating that [ɔ̃] and [ɑ̃] were more easily identified after occlusive consonants, but that neither manner of articulation influenced the perception of [ɛ̃]. The interaction Manner × Training group was not significant across the three vowels: [ɔ̃], $F(1, 38) = 0.18$, $p = .67$, $\eta_p^2 = .005$, [ɑ̃], $F(1, 38) = 2.69$, $p = .109$, $\eta_p^2 = .066$, and [ɛ̃], $F(1, 38) = 0.56$, $p = .456$, $\eta_p^2 = .015$, therefore consonantal context affected the perception of both training groups in a similar way in the AV test modality.

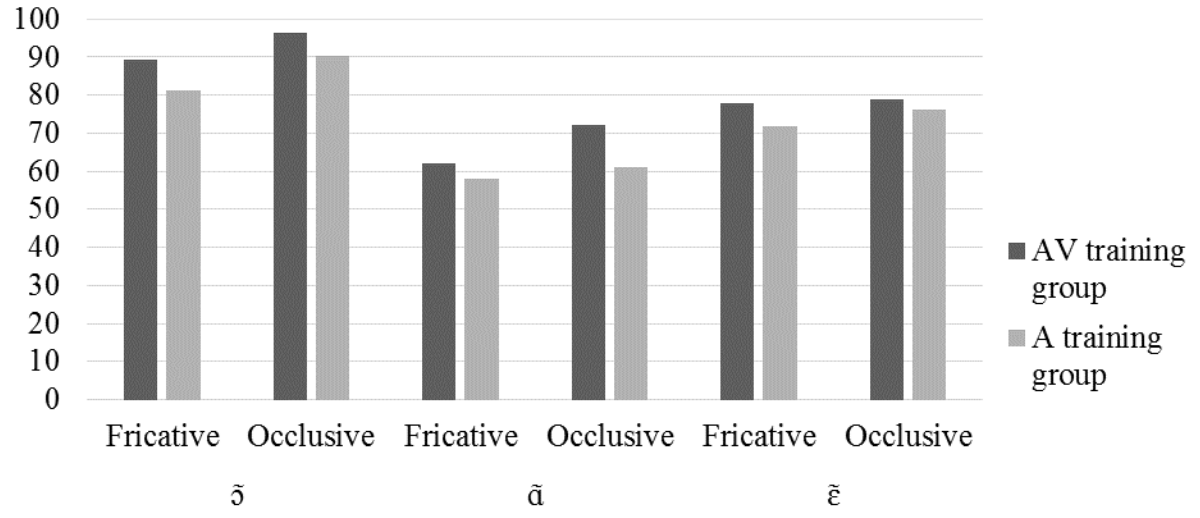

Figure 30. Percentage of accuracy score at the AV perception posttest according to manner of articulation of the initial consonant

In the Audio modality (Figure 31), there was also a significant main effect of Manner for [ɔ̃], $F(1, 38) = 20.52$, $p < .001$, $\eta_p^2 = .351$, and [ɑ̃], (1, 38) = 4.73, $p = .036$, $\eta_p^2 = .111$, but not for [ɛ̃], $F(1, 38) = 2.65$, $p = .111$, $\eta_p^2 = .065$, indicating that, like in the AV modality, [ɔ̃] and [ɑ̃] were more easily identified after occlusive consonants. The interaction Manner × Training group was again not significant across the three vowels: [ɔ̃], $F(1, 38) = 2.55$, $p = .110$, $\eta_p^2 = .063$, [ɑ̃], $F(1, 38) = 2.29$, $p = .138$, $\eta_p^2 = .057$, and [ɛ̃], $F(1, 38) = 0.23$, $p = .628$, $\eta_p^2 = .006$, indicating that the consonantal context affected the perception of both training groups in a similar way in the Audio modality.

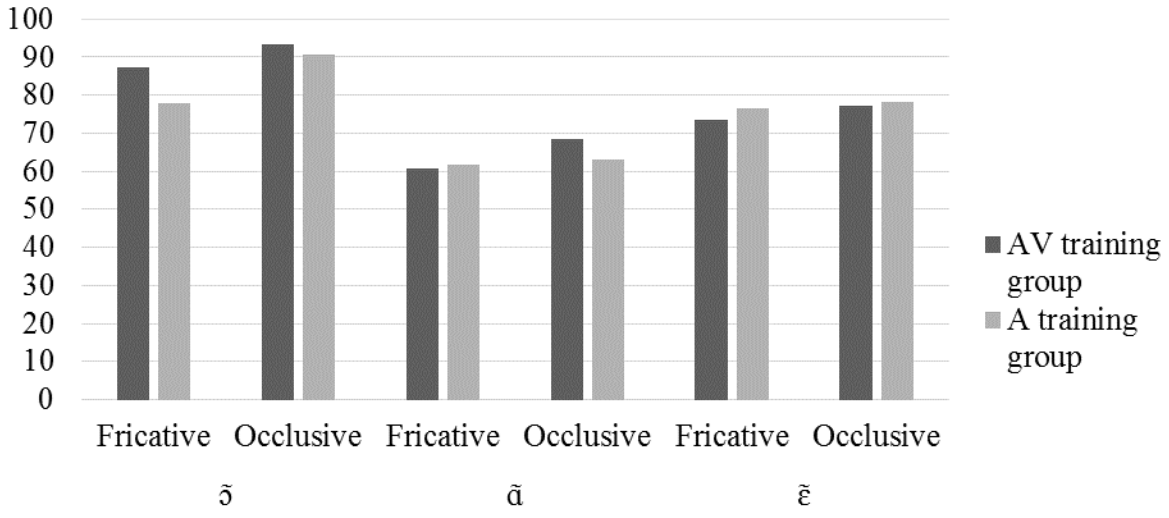

Figure 31. Percentage of accuracy score at the A-only perception posttest according to manner of articulation of the initial consonant

In the Visual-only modality (Figure 32), there was a significant effect of Manner for [ɔ̃], $F(1, 38) = 8.05$, $p = .007$, $\eta_p^2 = .175$, and [ɛ̃], $F(1, 38) = 5.08$, $p = .030$, $\eta_p^2 = .118$, but not for [ɑ̃], $F(1, 38) = 0.26$, $p = .614$, $\eta_p^2 = .007$. The interaction Manner × Training group was significant for [ɔ̃], $F(1, 38) = 5.07$, $p = .030$, $\eta_p^2 = .118$, [ɑ̃], $F(1, 38) = 4.56$, $p = .039$, $\eta_p^2 = .107$, but not for [ɛ̃], $F(1, 38) = 0.38$, $p = .542$, $\eta_p^2 = .010$, indicating that the two groups performed differently from each other for [ɔ̃] and [ɑ̃] according to the consonantal context. For the A-only training group, [ɔ̃] and [ɛ̃] were identified significantly more easily after occlusive consonants based on visual information only, whereas the preceding consonant did not affect the perception of [ɑ̃] positively or negatively. For the AV training group, perception ability was not affected by the consonantal context for [ɔ̃] and [ɛ̃], but the difference between fricative (45.8%) and occlusive (53.1%) approached significance ($p = .079$) for [ɑ̃].

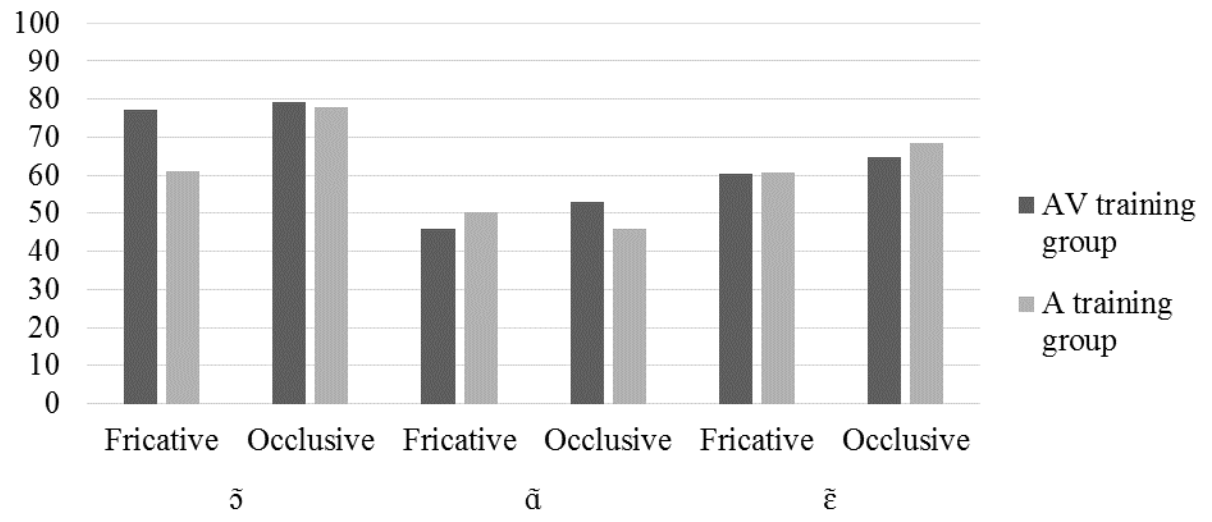

Figure 32. Percentage of accuracy score at the V-only perception posttest according to manner of articulation of the initial consonant

3.3.4 The effects of the final consonant

The same analyses conducted with the initial consonant were conducted with the final consonant. No significant effects were found for Labiality [$F_A(1, 38) = 3.26$, $p = .079$, $\eta_p^2 = .079$; $F_{AV}(1, 38) = 3.59$, $p = .071$, $\eta_p^2 = .086$; $F_V(1, 38) = 38.00$, $p = .855$, $\eta_p^2 = .001$]; Place of articulation [$F_A(5, 190) = 2.09$, $p = .068$, $\eta_p^2 = .052$; $F_{AV}(5, 190) = 1.04$, $p = .395$, $\eta_p^2 = .027$; $F_V(5, 190) = 0.34$, $p = .887$, $\eta_p^2 = .009$]; and Manner of articulation [$F_A(1, 38) = 1.56$, $p = .219$, $\eta_p^2 = .039$; $F_{AV}(1, 38) = 0.18$, $p = .669$, $\eta_p^2 = .005$; $F_V(1, 38) = 0.18$, $p = .668$, $\eta_p^2 = .005$]. In addition, none of the interactions Labiality × Training group, Place of articulation × Training group, and Manner of articulation × Training group were significant. The data therefore show that, for both training groups, final consonants did not influence accurate perception of the vowels.

3.4 Research question 4: Generalization

All the participants completed a generalization perception test following the perception posttest in order to see whether improvement gained during training would extend to novel

stimuli. Three types of stimuli were presented during the generalization test: (1) novel stimuli with the same CVC pattern as for the pretest and training, (2) CCVC stimuli with initial consonantal cluster [dʁ], and (3) CVCC stimuli with final consonantal cluster [dʁ]. Figure 33 illustrates the percentage of correct identification for the generalization test without distinguishing between the different types of stimuli presented. Similarly to the posttest results, the results of the control group were not significantly different from those of the pretest [A modality: $F(1, 18) = 0.32$, $p = .576$, $\eta_p^2 = .018$; AV modality: $F(1, 18) = 1.67$, $p = .211$, $\eta_p^2 = .085$; V modality: $F(1, 18) = 0.09$, $p = .77$, $\eta_p^2 = .005$]. The generalization scores for the control group also did not significantly differ from their results at the posttest, [A modality: $F(1, 18) = 0.00$, $p = .97$, $\eta_p^2 = .00$; AV: modality $F(1, 18) = 0.29$, $p = .59$, $\eta_p^2 = .016$; V modality: $F(1, 18) = 0.32$, $p = .57$, $\eta_p^2 = .018$].

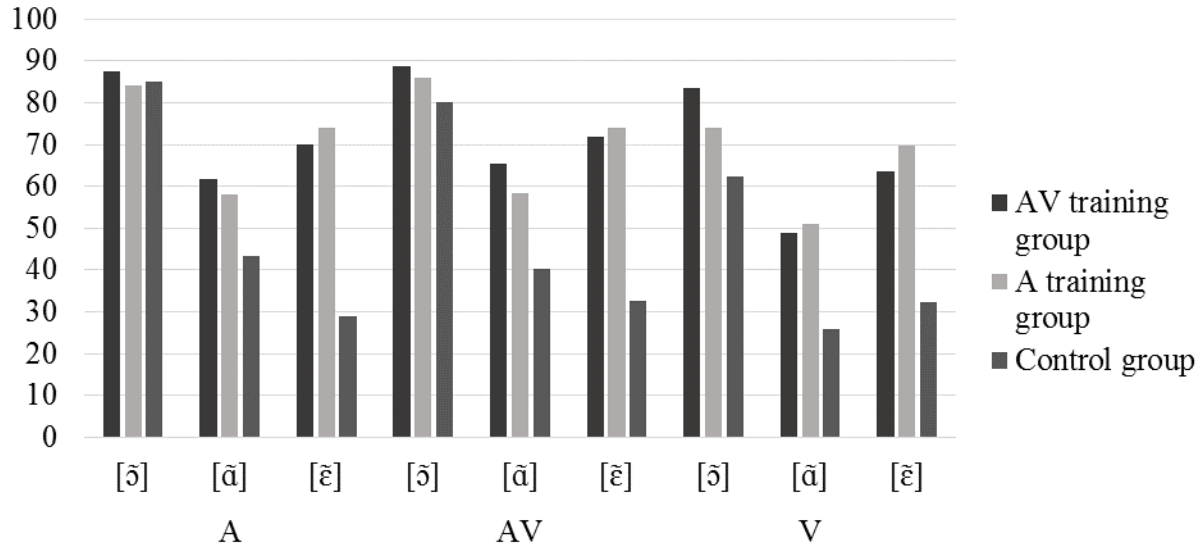

Figure 33. Mean percentage of correct identification for the generalization test

3.4.1 Comparison between the posttest and the generalization test

To assess the effects of training on generalization abilities in each testing modality (A, AV, and V), separate repeated-measures ANOVAs were conducted with the variables Time

(posttest and generalization test) and Training group (A and AV). Results show that for the A modality, Time had a significant effect [$F(1, 38) = 12.44$, $p = .001$, $\eta_p^2 = .247$], but not the interaction Time × Training group [$F(1, 38) = 0.39$, $p = .531$, $\eta_p^2 = .010$]. For the AV modality, there was also a significant effect of Time [$F(1, 38) = 5.21$, $p = .028$, $\eta_p^2 = .121$], and the interaction Time × Training group approached significance [$F(1, 38) = 3.35$, $p = .075$, $\eta_p^2 = .081$]. Finally, in the V modality, the main effect of Time was again significant [$F(1, 38) = 5.75$, $p = .021$, $\eta_p^2 = .131$], but the interaction Time × Training group was not [$F(1, 38) = 0.74$, $p = .740$, $\eta_p^2 = .019$]. Overall, the results suggest that, although there were significant changes between the posttest and the generalization test, both training groups performed similarly.

For the AV group, mean identification accuracy scores on the generalization test were 73% (A modality), 75% (AV modality), and 65% (V modality) (Table 18). Repeated-measures ANOVAs revealed that the generalization scores were significantly lower than the ones on the posttest in the A modality [$F(1, 19) = 8.15$, $p = .010$, $\eta_p^2 = .30$] and AV modality [$F(1, 19) = 6.66$, $p = .018$, $\eta_p^2 = .26$], but were similar to those in the V modality [$F(1, 19) = 0.97$, $p = .33$, $\eta_p^2 = .046$]. Analysis of the results per vowel (Table 19) revealed that performance decreased significantly from the posttest to the generalization test for [ɛ̃] in the A modality [$F(1, 19) = 4.90$, $p = .039$, $\eta_p^2 = .205$], and in the AV modality [$F(1, 19) = 4.94$, $p = .038$, $\eta_p^2 = .207$], but improved for [ɔ̃] in the V modality [$F(1, 19) = 5.45$, $p = .031$, $\eta_p^2 = .223$].

For the A-only training group, mean identification accuracy scores at the generalization test were 72% (A modality), 73% (AV modality), and 65% (V modality). Analysis revealed that there was a significant improvement at the generalization test in the V modality [$F(1, 19) = 6.73$, $p = .018$, $\eta_p^2 = .262$], a significant decrease in identification accuracy in the A modality [$F(1, 19) = 4.46$, $p = .048$, $\eta_p^2 = .19$], but no change between the posttest and generalization test in the AV

modality [$F(1, 19) = 0.14$, $p = .71$, $\eta_p^2 = .007$]. Analysis of the results per vowel showed that performance decreased for [ɛ̃] in the A modality [$F(1, 19) = 4.66$, $p = .04$, $\eta_p^2 = .197$], and that there was an increase approaching significance for [ɛ̃] in the V modality [$F(1, 19) = 4.19$, $p = .055$, $\eta_p^2 = .181$].

Table 18.

Mean percentage of correct identification at the posttest and generalization test

	Posttest			Generalization test		
	A	AV	V	A	AV	V
AV training group	77	79	63	73	75	65
A-only training group	75	73	61	72	73	65
Control group	53	53	42	53	51	40

Table 19.

Mean percentage of correct identification at the posttest and generalization test according to the vowel

		Posttest			Generalization test		
		AV training	A-only training	Control	AV training	A-only training	Control
A	[ɔ̃]	90	84	83	88	84	85
	[ɑ̃]	65	62	43	62	58	43
	[ɛ̃]	75	77	34	70	74	29
AV	[ɔ̃]	93	86	83	89	86	80
	[ɑ̃]	67	60	39	65	58	40
	[ɛ̃]	78	74	37	72	74	33
V	[ɔ̃]	78	70	62	84	74	62
	[ɑ̃]	49	48	25	49	51	26
	[ɛ̃]	63	65	38	63	70	32

3.4.2 Effect of modalities, vowels, and novel syllable structure

Recall that the novel stimuli presented during the generalization test consisted of three types of syllable structures: CV, CCVC, and CVCC where the consonantal cluster was [dʁ]. In order to see whether the syllable structure had an effect on identification accuracy, separate repeated-measures ANOVAs were conducted for each modality with Syllable structure (initial cluster, final cluster, no cluster) as within-subjects variables and Training group (A, AV, control) as between-subjects variable.

Results show that in the Audio modality (Figure 34), there was no significant effect of Syllable structure [$F(2, 112) = 0.01$, $p = .98$, $\eta_p^2 = .00$], but the interaction Syllable structure ×

Training group was significant [$F(4, 112) = 2.71$, $p = .033$, $\eta_p^2 = .088$], suggesting that the effect of syllable structure on perception varied between groups. Pairwise comparisons revealed that the differences were between the AV training group and the control group ($p = .002$) and between the A-only training group and the control group ($p = .004$). Further analysis showed that, although no syllable structure facilitated accurate vowel perception more than the other structures for the two training groups, initial consonantal cluster led to significantly higher vowel identification for the control group (57%) in comparison to the final cluster (51.9%), $p = .045$, and in comparison to the no cluster structure (51.6%), $p = .017$.

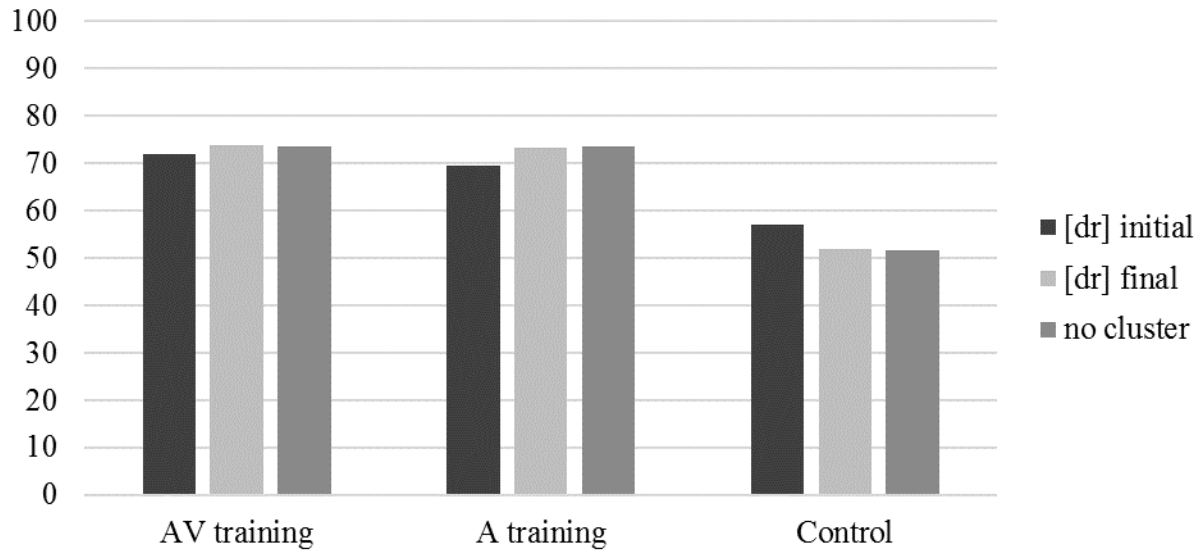

Figure 34. Percentage of correct identification in the A-only modality according to syllable structure

Results in the AV modality (Figure 35) revealed no significant effect of Syllable structure [$F(2, 112) = 0.82$, $p = .44$, $\eta_p^2 = .014$], and no significant interaction of Syllable structure × Training group [$F(4, 112) = 1.08$, $p = .36$, $\eta_p^2 = .037$], indicating that accurate AV perception did not differ between syllable structures and between groups.

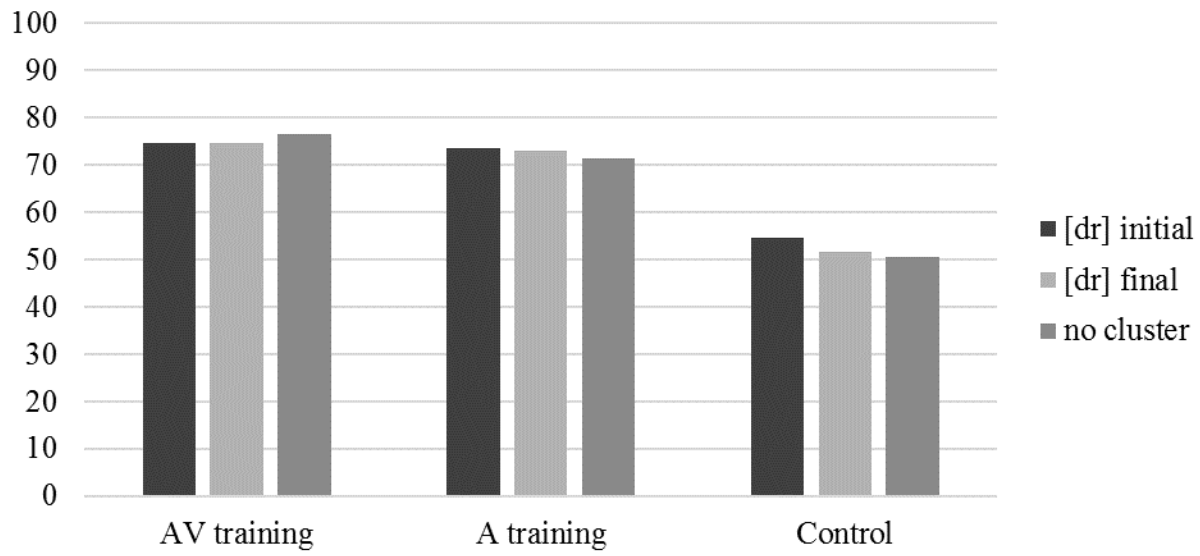

Figure 35. Percentage of correct identification in the AV modality according to syllable structure

In the Visual modality (Figure 36), Syllable structure was a significant main effect [$F(2, 112) = 12.20$, $p < .0001$, $\eta_p^2 = .179$], but the interaction Syllable structure × Training group was not significant [$F(4, 112) = 0.97$, $p = .426$, $\eta_p^2 = .034$], suggesting that accurate perception differed between syllable structures in the same way across the three groups. Pairwise comparisons revealed that accurate perception of the vowels was significantly higher for stimuli with initial consonant cluster (A-only training group: 70.1% and AV training group: 68.8%) than for final cluster (A-only training group: 63.1% and AV training group: 63.9%) and no cluster (A-only training group: 61.5% and AV training group: 63.3%) cluster, suggesting that initial consonant clusters facilitated accurate visual perception of the following vowel. The fact that syllable structure was not a significant factor for the control group also suggests that the facilitative effect of the initial consonantal cluster on vowel perception was due to training, although it remains unclear how the A-only training group benefited from visual information.

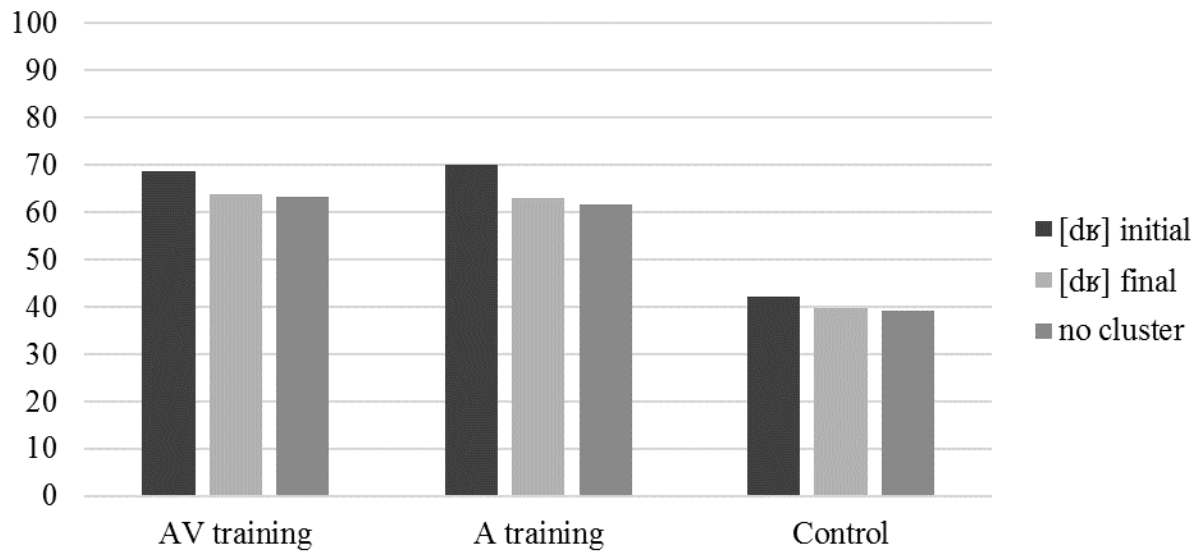

Figure 36. Percentage of correct identification in the V-only modality according to syllable structure

CHAPTER 4: DISCUSSION

This dissertation set out to answer the following research questions: (1) Does AV perceptual training lead to greater improvement in L2 perception of French nasal vowels than A-only training does? (2) Does AV perceptual training lead to greater improvement in L2 production of French nasal vowels than A-only training does? (3) Does perception accuracy vary in relation to consonantal context? (4) Is training generalizable to novel stimuli? This chapter summarizes the research results and discusses the findings in the light of other (audiovisual) speech perception studies. The following sections are organized according to research questions.

4.1 Research question 1: Perception

The main goal of this study was to examine the effects of two types of training on the perceptual learning of L2 French nasal vowels by American intermediate learners of French and to explore whether training that incorporates both visual and audio information led to more improvement than training with audio information only. Overall, the results presented in Chapter 3 show that the perceptual identification performance of both training groups increased significantly from pretest to posttest. The AV training group improved from 47.5% to 76.8% in the A modality, from 45.8% to 79.5% in the AV modality, and from 38.5% to 63.3% in the V modality. Similarly, the A-only training group improved from 46.3% to 74.6% in the A modality, from 42.8% to 73.1% in the AV modality, and from 37.4% to 60.8% in the V modality. On the other hand, the perception accuracy of the control group did not increase from pretest to posttest, demonstrating that any change occurring with the trainees was due to the training they received. The results are consistent with previous research showing that it is possible to successfully train L2 learners to modify their perception of audio and visual L2

speech contrasts (Bradlow et al., 1999; Hardison, 2003, 2005a; Hazan et al., 2005; Lively et al., 1993).

The results of the perception pretest and posttest also showed that there was an effect of vowel. At the pretest, the hierarchy of accurate perception in the AV and A modalities was [ɔ̃] > [ɑ̃] > [ɛ̃], with significant differences between the scores of each vowel. In the V modality, the hyper-rounded [ɔ̃] remained the best perceived vowel, but the unrounded [ɛ̃] was better perceived than the rounded [ɑ̃]. A possible explanation accounting for the low accuracy score for [ɑ̃] (26%) is that this vowel occupies the intermediate position on the continuum of labiality developed by Zerling (1989). It is neither unrounded nor hyper-rounded and it therefore lacks visual saliency in comparison to the two other nasal vowels. Observation of the confusion patterns (Table 9) also revealed that participants were consistent in their response patterns and in the errors they made. The vowel [ɔ̃] was perceived well in the A and AV modalities, but tended to be perceived as [ɛ̃] about 19% of the time. Because no visual information was provided in the A modality, the learners confused the two vowels solely based on the audio signal. The fact that similar results were obtained in the AV modality seems to suggest that learners attended more to the audio signal than to the visual signal when both were available. This would explain why they mistook the hyper-rounded [ɔ̃] for the unrounded [ɛ̃]—its opposite on the continuum of labiality. The results of the V modalities, however, showed that participants' selection of [ɛ̃] instead of [ɔ̃] increased to 30%, providing evidence that they did not have accurate representations of the vowel [ɛ̃] and how it is produced. This is confirmed by the low results obtained when [ɛ̃] was presented. Participants tended to overwhelmingly choose the rounded [ɑ̃] instead of the accurate [ɛ̃] (i.e., above 60% in all modalities). Finally, the results for [ɑ̃] also showed that the learners' perceptual representations were not accurate. In the A-only modality, although 45% of the

responses were correct, participants also hesitated between [ɔ̃] (37%) and [ɛ̃] (18%). On the other hand, when visual information was added and even more so when only visual information was available, [ɔ̃] was the first response choice. This suggests that L2 learners tended to associate roundedness with [ɔ̃].

At the posttest, the hierarchy of accurate perception remained unchanged for the control group: [ɔ̃] was still consistently better perceived across the three modalities, [ɑ̃] was better perceived than [ɛ̃] in the A and AV modalities, but less accurately perceived than [ɛ̃] in the V modality. Training was shown to be particularly beneficial for [ɛ̃] and, as a result, the hierarchy of accurate identification for the participants in both training groups changed: [ɔ̃] remained the best perceived vowel, with scores reaching above 85% in the A and AV modalities and above 70% in the V modality, and [ɛ̃] became the second best perceived vowel across all modalities.

Regarding the confusion patterns (Table 13), results show that, after training, participants did not confuse [ɔ̃] and [ɛ̃]—the two vowels at the end of the continuum of labiality—as much as they did at the pretest. Across the three modalities, although the effects were more marked in the V modality, the intermediate vowel [ɑ̃] was the option taken when participants did not select the correct answer. Despite the progress in perception for participants in the two training groups, [ɑ̃] still remained challenging and was more confused with [ɔ̃] than it was with [ɛ̃], especially when only visual input was provided, indicating that L2 learners did realize that [ɑ̃] is a rounded vowel. As previous studies have noted (e.g., Hardison, 2003; Hazan et al., 2006; Wang et al., 2009), difficulties in perceiving L2 visual contrasts may be attributed to the influence of the visual cue inventory of the native language. Because French has more rounded vowels than American English has and because American English does not have hyper-rounded vowels, American L2 learners of French might tend to assimilate French rounded and hyper-rounded

vowels to one single category: rounded. Therefore, they need to learn to distinguish between two degrees of roundedness, and to learn to associate these two degrees of roundedness to corresponding L2 phonemes and visemes in order to establish new L2 categories.

Despite improvement made by the two training groups in the current study, it is important to note that, contrary to numerous previous studies in AV speech perception, training two modalities (A and V) simultaneously was not superior in improving perceptual accuracy to training only one. In particular, learners trained audiovisually did not improve their perception of the vowels in the V modality significantly more than those trained auditorily. In addition, analysis of the accuracy scores during each of the six training sessions (Figure 11) showed that both training groups improved in similar fashions. No group outperformed the other in terms of accuracy scores, both groups consistently improved from session to session and stopped improving after the fifth session. The possibility that a longer training program (i.e., either involving longer duration per session or additional sessions) would have led to significant differences between the two training groups seems therefore ruled out. In addition, even if we consider that the stop in improvement was just a temporary plateau—since the groups still improved numerically—there is no indication that suggests that the AV training group would have started to improve more than the A training group after the sixth training session.

The lack of AV effects in previous studies has sometimes been explained due to the lack of visual saliency between the contrasts under investigation and the mapping of L2 phonemes to L1 categories. For instance, contrary to Hardison (2003), Hazan and her colleagues (2005) did not find significant differences between the AV and A-only training groups when tested on the English /r/-/l/ contrast. Similarly, Spanish participants receiving AV training did not perform better than A-only trainees when asked to identify the English /v/-/b/ and /ð/-/d/ contrasts

(Ortega-Llebaria et al., 2001). Although the reason is still unclear why AV training was not superior to A-only training, the authors proposed that their Spanish participants "may have learnt to disregard certain visual cues to place/manner in their L1" (p. 152). In this scenario, the distinction between the pairs of visemes would, therefore, not be meaningful, resulting in participants perceiving the two contrastive visemes as one single viseme. In the current study, the situation is novel and not directly comparable as it involves the distinction between, not two consonants, but three vowels. Nonetheless, the differences between three nasal vowels are visually and auditorily salient. From a visual perspective, they belong to three distinct viseme categories[5], as demonstrated by research on lipreading and visual speech synthesis (Benoît, Lallouache, Mohamadi, Tseva, & Abry, 1991; Zerling, 1990). In a previous study (Inceoglu, 2011), I tested native French speakers on their perceptions of the three nasal vowels in the same three modalities of presentation (A, AV, V) as used in the current study. My results showed that participants identified the three vowels accurately more than 98% in the AV modality and more than 99% in the A-only modality. As for the V modality, native speakers correctly identified [ɔ̃] 97% of the time, [ɛ̃] 94% of the time, and [ɑ̃] 70% of the time. The fact that [ɑ̃] was visually the least accurately perceived vowel by native speakers is similar to the results of the current study. Nevertheless, the score obtained by both the native speakers and the participants in the two training groups were above chance level. In terms of confusion pattern (Table 13), native speakers tended to confuse the rounded [ɑ̃] with the hyper-rounded [ɔ̃] 27% of the time and they very rarely picked the unrounded vowel [ɛ̃] (3%). The L2 learners confused [ɑ̃] with [ɔ̃] about 38% of the time, but also picked [ɛ̃] 13% of the time, suggesting that they did not only confuse

[5] French vowels are classified into seven viseme categories

rounding and hyper-rounding, like the natives did, but also tended to perceive [ɑ̃] as an unrounded vowel.

Another possible explanation for the lack of significant differences between the two types of training in the current study is the fact that attending to two types of information (audio and visual) might have caused an overload of the cognitive processes involved. Contrary to previous training studies investigating two (consonantal) contrasts, exposing L2 learners to three (vocalic) contrasts might have further increased the cognitive load of the task and diminished reliance on one type of information, particularly in this case the visual information. Participants in the AV training group might have found the audio information more helpful for contrasting the three vowels than the visual information, which would explain why their performances were similar to those of the A-only training group.

Studies in SLA report strong evidence of individual variability due to factors such as personality, aptitude, motivation, learning styles and learning strategies (Dörnyei & Skehan, 2003). In addition to these personal characteristics, there is also great variability in learners' lipreading skills (Demorest, Bernstein, & DeHaven, 1996; Summerfield, 1992) and how individuals integrate visual and auditory information (Grant & Seitz, 1998). As previously noted by Lisker and Rossi (1992) individual differences in lipreading greatly influence AV speech perception as preferences for an attentional focus on visual or audio information differ across individuals. In their study, the authors noticed that some participants focused mostly on the auditory signal, while others relied heavily on lipreading. Other studies have also highlighted the variability in lipreading behaviors by showing that men relied less on lipreading than women (Daly, Bench, & Chappell, 1996; Johnson, Hicks, Goldberg, & Myslobodsky, 1988). Although analysis of individual variations will be the subject of further research, some remarks can be

made at this time. First, the data from the current study do not indicate that there were any differences in AV perception between women and men. However, the analysis is limited due to the unbalance in the gender of the participants. Recall that 43 female and 17 male learners of French participated in the study. Although a more balanced distribution would have been ideal, this distribution actually represents classroom distribution and is therefore a good representative sample of the student population. More important is the learning style of the participants. Research on individual differences suggests that some individuals are more visual, while some others are more aural, or—less relevant to the current topic—tactile (Oxford & Anderson, 1995; Reid, 1995). Based on these preferences, it is therefore possible that some learners in the current study benefited more from the training condition they were assigned to than others did. Some poor lip readers in the AV training group might not have taken advantage of the visual information, whereas the learning of some visual learners assigned to the A-only training group might have been hindered by the lack of visual information. Conversely, because of variability in learning styles, it is also plausible that the visual cues in the AV modality acted as distractors for aural learners.

Another individual factor to take into account is the difference in experience with the L2. In a study investigating the influence of visual cues on the perception of the English /r/-/l/ contrast by Japanese learners in the United States (ESL), Hardison (2003) found that AV training was more effective than A-only training. On the other hand, in a similar study with Japanese learners of English in Japan (EFL), Hazan et al. (2005) did not find significant differences between the two types of training. A possible explanation is that the degree of exposure to the L2 might have an effect on the availability of visual cues, the richness of visual cues due to a wide range of native talkers, and the learning of these cues. More exposure to the target language, in

the case of second language settings, might also affect the motivation of the L2 learners and therefore their readiness to attend to audio and visual information during training. Similarly to Hazan et al. (2005), the current study was conducted in a foreign language setting and participants had little contact with the target language outside of class. None had spent more than two weeks in a Francophone country and all reported similar exposure to French outside of the classroom.

4.2 Research question 2: Production

The second question investigated the relationship between perception and production. Previous auditory training studies have shown that gain made during perceptual training can be transferred to gain in production, even when participants did not receive any specific training on how to produce the words (e.g., Bradlow et al., 1999, 1997; Lambacher et al., 2005; Lopez-Soto & Kewley-Port, 2009; Wang, Jongman, & Sereno, 2003). Although much less research has been conducted in AV training studies, findings so far have demonstrated that AV perceptual training leads to greater improvement in production than A-only training (Hardison, 2003; Hazan et al., 2005).

Overall, the results of the current study are consistent with these findings, but the two types of analysis used to assess the production of the L2 learners (i.e., identification task and quality rating task) revealed some differences. The results of the identification task showed that both training groups significantly improved from the pretest to the posttest, while the improvement of the participants in the control group was not statistically significant. The significant changes in the production of the trainees at the posttest and the lack of changes in the production of the control group are evidence that a transfer from perceptual training to production occurred. In addition, in line with Hardison (2003) and Hazan (2005), the production

of the AV training group improved significantly more than the production of the A-only training group did, suggesting that participants in the AV training group might have benefited and learned from the provision of visual information even if they did not exploit this information during the perception posttest. Results of the identification task also revealed that at the pretest the L2 learners' production of [ɛ̃] was better identified by the native raters than that of [ɔ̃], while [ɑ̃] received the lowest accuracy rating. The pattern remained the same at the posttest, with [ɛ̃] showing the greatest improvement (reaching about 90% accuracy for both training groups) and [ɑ̃] showing little improvement (increasing from about 55% to 59%). The results of accurate production based on the quality rating task showed similar patterns: [ɛ̃] was the best rated vowel, followed by [ɔ̃] and [ɑ̃]. However, the results of the vowel quality rating task did not point to the same improvement as the results of the identification task. Based on the quality rating task all groups improved their production of the nasal vowels. Nonetheless, consistent with previous studies and with the results of the identification rating, the improvement in production of the AV group was significantly greater than that of the A-only group, and the improvement of the control group was significantly lower than that of the two training groups. In sum, the results of the two assessment tasks led to converging results in favor of the AV training efficiency. The identification task provided a direct, segment-specific assessment of the nasal vowel production, while the quality rating task shed light on a more specific evaluation of the participants' pronunciation.

Researchers have used various techniques to rate the production of L2 learners, sometimes combining several rating tasks (e.g., Bradlow et al., 1999; Hazan et al., 2005) and other times using only one (e.g., Hardison, 2003; Wang et al., 2003). In the current study, I used a quality rating task and a minimal identification task, although contrary to Hazan et al. (2005)

the latter task did not involve a minimal pair but a triad. The fact that the native French raters had to make a choice out of three options considerably complicated the task. The chances for the learners' productions to be accurately rated by the raters was reduced, the cognitive demand for the raters was increased, and the total accuracy scores might be lower than if only two sounds had been contrasted. An additional drawback of a forced identification task is that the raters do not have the possibility to select a "none of these sounds" option. This leads to the risk of raters choosing one option by default or lack of better choice. Quality rating tasks do not have this problem, but they can sometimes be influenced by rater biases (Hoyt, 2000). For instance, the raters in this study seldom used the lower numbers on the seven-point Likert scale which led to relatively high quality scores. This tendency to inflate scores is, however, not problematic for the current study as this propensity was consistent throughout the rating task and since pretest and posttest tokens were randomized. Other rating methods exist, but they were deemed less informative and more time consuming for the analysis of the 8640 production tokens (i.e., 108 stimuli × two times × 40 participants). For instance, Bradlow and her colleagues (1999) used a preference rating task and an open-set transcription task. In the former, the raters directly compared the pretest and posttest production of each word and assigned a grade on a Likert scale to compare the two productions. If the production of the first token was much better than the production of the second token, the raters assigned a one. If the opposite happened, raters assigned a seven, and if both productions were similar, raters gave a four. In the open-set transcription task, raters were asked to type the word they heard in order to provide a strict measure of overall intelligibility.

The current findings suggest that there is some link between L2 speech perception and production and, therefore, provide evidence for the claims made by some of the major speech

theories that speech perception and production are interdependent. A widely supported hypothesis is that accurate perception of L2 phones precedes accurate production (Flege et al., 1997; Flege, 1987, 1995) and, therefore, difficulty in distinguishing sounds causes difficulty in production (Bohn & Flege, 1992). The Motor Theory (Liberman & Mattingly, 1985) and the Direct Theory approach (Best, 1995) agree with this interdependence of perception and production, but argue that improvements in speech perception should be simultaneously accompanied by improvements in production based on the assumption that L2 sounds are perceived directly in relation to the articulatory gestures of the speaker. As shown in the current study, the improvement in pronunciation for the AV training group was greater than the improvement of the A-only training group, but both groups did not differ at the perception posttest. This suggests that the development of the AV trainees' production abilities did not occur linearly with the development of their perception abilities. Although inconsistent with the two views mentioned above, previous studies have also reported that L2 perception and production do not always follow parallel developments. In a perceptual training study on English vowels by Chinese ESL learners, Wang (2002) found that training effectively helped improve learners' perception abilities, but did not transfer to production. On the other hand, researchers have also reported that accurate production of certain L2 sounds precedes the perception of those sounds (Borden et al., 1983; Gass, 1984; Goto, 1971; Sheldon & Strange, 1982; Smith, 2001; Zampini & Green, 2001). In a study investigating voicing by Spanish ESL learners, Zampini and Green (2001) found that learners showed a short voice onset time in production before they were able to perceive it. Similar findings were noted with Japanese learners of English producing the English /r-l/ contrast before perceiving it accurately (Sheldon & Strange, 1982; Smith, 2001). A study on the perception and production of vowels by Korean second language learners found that

perception did precede production for most of the learners, but highlighted some individual differences that are worth mentioning here (Baker & Trofimovich, 2006). The authors reported that the perception/production relationship seemed to depend on the length of residence of the L2 learners, and noted that "perception and production may be aligned at initial and more advanced stages of L2 learning. However, in the intermediate stages of L2 learning (where presumably most of the learning occurs), perception and production skills are misaligned" (p. 246). These findings are particularly relevant to this dissertation as the current participants were recruited in intermediate French classes. Their proficiency and their exposure to the L2 might therefore be reasons that accounted for the fact that their production seemed to precede their perception of the French nasal vowels. Finally, a complementary possibility is that the perception task invited explicit reasoning more than the production task did, and that for motor tasks conscious knowledge may sometimes interfere with performance.

4.3 Research question 3: Consonantal context

The third research question addressed the issue of the consonantal context (i.e., preceding and following consonants) and its possible influence on the perception of the vowels. The data from the present study revealed that the following consonant did not affect vowel perception, but the preceding consonant did. Perception accuracy was higher when the initial consonant was non labial (vs. labial) and when it was an occlusive (vs. fricative). More specifically, higher identification scores were obtained with velars and bilabials, while palatals led to the lowest accuracy scores. A possible explanation for the observed differences regarding velar consonants is that they are articulated with the back part of the tongue and are therefore not visually salient. Contrary to labial consonants, and especially palatals, the movement of the lips involved while producing [k] and [g] sounds does not interfere with the labial movement for the vowel. This,

therefore, increases the saliency of the vowels by reducing coarticulatory influences. The reason why the bilabial context also produced high accuracy scores cannot, however, be explained with the same arguments. Nevertheless, these results support what Montgomery, Walden, and Prosek (1987) reported in their investigation of the effects of consonantal context on vowel lipreading. They found that for one of the two talkers used in the experiment the velar [gVg] and bilabial [pVp - bVb] contexts led to higher accuracy scores, while the palatal and labiodental contexts produced lower identification accuracy scores—just like in the current study. An important aspect to keep in mind is that Montgomery et al. (1987) found a significant interaction between the phonetic context and their two talkers, but because only one talker was used in the current experiment, no further investigation of talker effect can be made.

The comparison between the two manners of articulation offered results less ambiguous than those for the place of articulation. Findings revealed that the vowels benefited greatly from the occlusive context as higher accuracy scores were obtained with velar, bilabial, and dental consonantal contexts. On the other hand, contexts with a fricative consonant produced significant decreases in perception accuracy. Montgomery et al. (1987) have found similar results for English lax vowels and have suggested that "stops, with their more rapid opening and closing gestures, give the viewer good information on vowel duration, whereas the continuant consonants with their more gradual transitions tend to obscure the vowel onsets and terminations and make visible vowel duration more variable" (p. 57). Although French nasal vowels are not comparable to English lax vowels, it seems probable that the duration of the initial consonant had an influence on the perception of the vowel. Shorter consonantal gestures, in the case of occlusives, might have facilitated the identification of the vowel even when the occlusives had

highly visible labial components (i.e., bilabials). On the other hand, vowel recognition might have been affected by the long duration information available in fricative contexts.

Overall, the two training groups appear to have been affected by the aforementioned phonetic contexts in the same way. This indicates that the lack of significant differences at the perception posttest between the two groups was not due to the consonantal context and that both groups improved their perceptual accuracy equally. In the AV and A-only testing modalities, both groups recognized [ɔ̃] and [ɑ̃] better in occlusive contexts than they did in fricative contexts, while [ɛ̃] was not affected by any of the contexts. In the V-only modality, the A-only training group identified [ɔ̃] and [ɛ̃] better in occlusive contexts, but for the AV training group the difference between occlusives and fricatives was not significant (albeit approaching significance).

In conclusion, in line with previous studies (Benoît et al., 1994; Gottfried, 1984; Hardison, 2003; Montgomery et al., 1987; Owens & Blazek, 1985; Strange et al., 2001), the current data suggest that the perception of a phoneme or viseme (either vowel or consonant) is not based on its individual characteristics, but is also influenced by the phonetic context. This has obvious implications for how to design stimuli in auditory and AV perception studies and provides further support in favor of the use of high variability materials as different phonetic contexts have different effects.

4.4 Research question 4: Generalization

The final research question investigated whether participants could transfer learning from stimuli presented during perceptual training to new stimuli. Researchers have noted that the goal of a successful training is to demonstrate that the learning effects can be generalized. Because previous studies have already shown that training can be generalizable to novel stimuli produced

by a novel talker (Bradlow et al., 1997; Hardison, 1996, 2003; Lively et al., 1994; Logan et al., 1991; Nishi & Kewley-Port, 2007; Pruitt et al., 2006), the current study did not further investigate the issue of novel voice, but instead focused on the generalization to novel stimuli. Three types of novel stimuli were presented: (1) novel stimuli with the same consonantal contexts, (2) novel stimuli with the initial consonantal cluster [dʁ], and (3) novel stimuli with the final consonantal cluster [dʁ]. Overall, participants in the two training groups demonstrated comparable performance at the posttest and generalization test. In the A-only modality, both groups performed significantly better at the posttest than at the generalization test and their identification accuracy significantly decreased for [ɛ̃] at the generalization test. In the AV modality, the performance of the AV training group significantly decreased, but there was no significant difference between the posttest and the generalization test for the A-only training group. More importantly, no differences were observed between the three types of stimuli in the AV and A-only modalities. In the V modality, however, both training groups performed significantly better in the initial cluster context than in the final cluster and the no cluster contexts. In addition, the performance of the A-only training group was significantly better at the generalization test, and although the AV training group's scores were similar between the two tests, their performance with [ɔ̃] increased at the generalization test. In sum, participants were able to extend the effects of their training to novel stimuli with a novel consonantal context and even improved in the V-only modality compared to how they did on the posttest. A possible explanation for the higher scores for the consonantal cluster [dʁ] in the V-only modality is that the longer duration of the cluster provided more time for the participants to perceive the vowel. This was facilitated by the fact that since both the dental [d] and the dorso-uvular [ʁ] are non labial consonants, the consonantal cluster did not affect the labiality of the vowel. Thus, it seems

that the initial consonantal cluster enhanced the saliency of the vowels and permitted the participants to increase their identification performances.

CHAPTER 5: CONCLUSION

This final chapter is divided into three parts. First, I will summarize the results of the four research questions that guided this study. Next, I will present the practical and pedagogical implications of the study. Finally, I will conclude with a brief discussion of limitations of the study, in addition to suggestions for future research.

5.1 Summary of the findings

The findings of this study can be summarized as follows. The two types of training (AV and A) used in this study had beneficial effects on perception accuracy of the French nasal vowels, while the control group, who did not receive training, did not show signs of improvement. Nevertheless, contrary to previous AV training studies on consonants, the present study did not find significant differences in the type of training used. Training that presented both audio and visual information did not lead to greater improvement in perception than training with audio information only. However, the current study found that, although both training groups improved on the production posttest, the AV training group improved their production accuracy significantly more than the A-only training group did. This suggests that the AV trainees might have used the visual information provided during training and transferred this information to improve their production abilities. This study also shows that perception was significantly more accurate when the initial consonant was non-labial (vs. labial) and an occlusive (vs. fricative). Finally, in order for training to be deemed effective, it is important to test the generalization of learning (Logan et al., 1991). The findings of the generalization test administered in this study are compatible with previous studies, namely by showing that participants were able to transfer the benefits of the training to novel stimuli.

5.2 Practical and pedagogical implications

Previous AV perceptual training studies have focused on L2 consonant contrasts and much remains to be explored regarding multimodal training of L2 vowels. This study has contributed to research in AV speech perception by providing empirical data showing that training which involved both audio and visual information did not lead to greater improvement in perception than auditory-only training. Nevertheless, the results of this study have shown that training was beneficial as it led to improvement in L2 speech perception accuracy. In addition, evidence showing that perceptual training can result in improvement in pronunciation, especially when training involves both audio and visual information, provides implications for language learning. The teaching of pronunciation is often marginalized in the language classroom (Derwing & Munro, 2005) and perceptual trainings, such as the one presented in this dissertation, have the potential to enhance pronunciation teaching and learning by developing L2 students' awareness of how sounds are produced without taking time away from the classroom. In addition, the present audiovisual perceptual training improved the learners' ability to perceive and produce French nasal vowels without providing explicit phonetics instruction on how the sounds are produced. It would, nevertheless, be interesting to investigate whether AV training can benefit from supplemental explicit instruction (i.e., asking participants to focus their attention to the speaker's lips and/or teaching them about the different degrees of labiality). Although the setting and procedures are not comparable, a recent classroom-based study by Kissling (2013) investigating production demonstrated that explicit phonetic instruction did not provide any advantage over a control condition that consisted of focused listening with dictation, but no explicit instruction.

The implications of this study are particularly relevant to the area of computer-assisted language learning (CALL). First and foremost, it reinforces the fact that speech is a multimodal experience, and that pedagogical tools should move away from auditory based materials to incorporate more audiovisual materials. Many websites that accompany recent textbooks do not feature audiovisual material although this would be very easy with current technology. In addition, some of the problems with current pronunciation software and online tools are that they do not always provide adequate feedback and sometimes fail to give accurate rating of the learners' pronunciation. On the other hand, perception trainings (either auditory only or audiovisual) are more reliable in terms of feedback, do not involve rating, and have shown to lead to improvements in both perception and production (Bradlow et al., 1999, 1997; Hardison, 2003; Hazan et al., 2005; Lambacher et al., 2005; Lopez-Soto & Kewley-Port, 2009; Y. Wang et al., 2003).

5.3 Limitations and further research

One limitation of this study is that the stimuli used for the pretest, posttest, training, and generalization test were a mix of pseudo words and real words. As mentioned in the Method section, because the experiment involved triads of stimuli with the three French nasal vowels, it was not possible to either have only real words or only pseudo words. Due to time limitation, word familiarity was not investigated in this study. However, because a pilot study with similar participants (i.e., same institution, same semester, similar background) showed that word familiarity did not affect (AV) perception of nasal vowels (Inceoglu, 2012), it is expected that word familiarity was not a confounding variable in this experiment either.

Another limitation that should be acknowledged is that the current experiment only tested words in isolation, and therefore no generalization can be made regarding the transfer from

words in isolation to connected speech. A previous study by Pereira (2012) showed that although multimodal training on English vowels by Spanish speakers was successful at the word-level, improvement was not transferred to processing of words in sentences.

Finally, the set of stimuli tested in this experiment was limited to the three French nasal vowels. The rationale for using nasal vowels was that (1) they are often problematic for L2 learners and therefore any improvement due to training would have pedagogical implications, (2) their contrast is visually salient and is based on lip rounding/spreading, which has not been investigated before. It may be that different results would be obtained with vowel pairs with different visual contrasts, such as vowels differing in terms of mouth opening. Based on these facts, and because few studies have investigated the AV perception of L2 vowels, no generalization should be made regarding the efficiency of AV training versus A-only training on vowels.

Based on the limitations mentioned above, future studies should expand the investigation of multimodal training by looking at more vowel contrasts. For example, comparably to Navarra and Soto-Faraco (2007) who investigated the effects of visual speech information on the perception of the L2 Catalan [e-ɛ] contrast, experiments could be conducted with learners of French as French also possesses this contrast. In any case, more research needs to be conducted with a greater range of vowel contrasts, a greater range of target languages and native languages, and a greater range of L2 proficiency.

Finally, future research should further expand the effect of learning styles in AV speech perception study to explore whether the experimental group participants are assigned to match their preferred way of learning. It remains to be seen whether visual participants would benefit more from AV training than aural participants assigned to the same training group. Conversely,

some visual participants might not find Audio-only training efficient as they tend to prefer learning with visual information.

APPENDICES

Appendix A: Background questionnaire

1. Age: _____
2. Gender: Male ☐ Female ☐
3. Mother tongue (First language):_____
4. Year in college:

Freshman ☐ Sophomore ☐ Junior ☐ Senior ☐ Graduate ☐ Other_____

5a. Are you a French major? Yes ☐ No ☐ Are you a French minor? Yes ☐ No ☐

5b. Are you continuing with French next semester? Yes ☐ No ☐

6. List the French language university classes you have taken.

Course (number and title)	When?	Required? (Yes or No)

7. Please circle your proficiency level for French in the following areas.

	Beginning					Advanced
Reading	1	2	3	4	5	6
Writing	1	2	3	4	5	6
Listening	1	2	3	4	5	6
Speaking	1	2	3	4	5	6

9. At what age did you begin studying French? _____
10. How long have you been studying French? _____
11. Have you ever visited a French-speaking country? Yes ☐ No ☐

If yes, where and for how long?

Location	Length of visit	Reason (study abroad, tourism..)

12. Outside of class, how many hours per week do you spend using French? _____

Watching movies/TV in French? _____

Listening to the radio/songs/news in French?_____

13. Do you have family members who speak French? Yes ☐ No ☐

If so, who (e.g., parents, grandparents, etc.)?_____

14. Please list any other languages that you have previously studied:

Language	Length of study

15. Do you have hearing problems? Yes ☐ No ☐

16. Do you have vision problems? Yes ☐ No ☐

MERCI!

Appendix B: List of generalization stimuli

Table 20.

Stimuli with [dʁ] as initial consonantal cluster (n = 36)

		[ɔ̃]	[ɑ̃]	[ɛ̃]
Final consonant	[p]	dʁɔ̃p	dʁɑ̃p	dʁɛ̃p
	[b]	dʁɔ̃b	dʁɑ̃b	dʁɛ̃b
	[t]	dʁɔ̃t	dʁɑ̃t	dʁɛ̃t
	[d]	dʁɔ̃d	dʁɑ̃d	dʁɛ̃d
	[k]	dʁɔ̃k	dʁɑ̃k	dʁɛ̃k
	[g]	dʁɔ̃g	dʁɑ̃g	dʁɛ̃g
	[s]	dʁɔ̃s	dʁɑ̃s	dʁɛ̃s
	[z]	dʁɔ̃z	dʁɑ̃z	dʁɛ̃z
	[f]	dʁɔ̃f	dʁɑ̃f	dʁɛ̃f
	[v]	dʁɔ̃v	dʁɑ̃v	dʁɛ̃v
	[ʃ]	dʁɔ̃ʃ	dʁɑ̃ʃ	dʁɛ̃ʃ
	[ʒ]	dʁɔ̃ʒ	dʁɑ̃ʒ	dʁɛ̃ʒ

Table 21.

Stimuli with [dʁ] as final consonantal cluster (n = 36)

		[ɔ̃]	[ɑ̃]	[ɛ̃]
Initial consonant	[p]	pɔ̃dʁ	pɑ̃dʁ	pɛ̃dʁ
	[b]	bɔ̃dʁ	bɑ̃dʁ	bɛ̃dʁ
	[t]	tɔ̃dʁ	tɑ̃dʁ	tɛ̃dʁ
	[d]	dɔ̃dʁ	dɑ̃dʁ	dɛ̃dʁ
	[k]	kɔ̃dʁ	kɑ̃dʁ	kɛ̃dʁ
	[g]	gɔ̃dʁ	gɑ̃dʁ	gɛ̃dʁ
	[s]	sɔ̃dʁ	sɑ̃dʁ	sɛ̃dʁ
	[z]	zɔ̃dʁ	zɑ̃dʁ	zɛ̃dʁ
	[f]	fɔ̃dʁ	fɑ̃dʁ	fɛ̃dʁ
	[v]	vɔ̃dʁ	vɑ̃dʁ	vɛ̃dʁ
	[ʃ]	ʃɔ̃dʁ	ʃɑ̃dʁ	ʃɛ̃dʁ
	[ʒ]	ʒɔ̃dʁ	ʒɑ̃dʁ	ʒɛ̃dʁ

Table 22.

Stimuli with CVC structure (n = 36)

	Occlusive-occlusive	Fricative-Fricative	Occlusive-Fricative	Fricative-Occlusive
[ɔ̃]	dɔ̃p	ʃɔ̃ʃ	tɔ̃f	sɔ̃b
[ɔ̃]	bɔ̃t	vɔ̃s	pɔ̃z	fɔ̃t
[ɔ̃]	gɔ̃g	zɔ̃f	kɔ̃ʃ	ʒɔ̃g
[ɑ̃]	pɑ̃b	sɑ̃z	kɑ̃v	sɑ̃p
[ɑ̃]	kɑ̃d	ʒɑ̃v	tɑ̃ʒ	ʒɑ̃d
[ɑ̃]	tɑ̃g	fɑ̃ʃ	pɑ̃s	fɑ̃k
[ɛ̃]	dɛ̃k	zɛ̃s	gɛ̃z	zɛ̃t
[ɛ̃]	gɛ̃d	ʃɛ̃v	bɛ̃f	vɛ̃k
[ɛ̃]	bɛ̃p	vɛ̃ʒ	dɛ̃ʒ	ʃɛ̃b

REFERENCES

REFERENCES

Alm, M., Behne, D. M., Wang, Y., & Eg, R. (2009). Audio-visual identification of place of articulation and voicing in white and babble noise. *The Journal of the Acoustical Society of America*, *126*(1), 377–387.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*(02), 339–355.

Assmann, P. F., Nearey, T. M., & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *The Journal of the Acoustical Society of America*, *71*(4), 975–989.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels : The role of individual differences. *International Review of Applied Linguistics in Language Teaching*, *44*(3), 231–250.

Battye, A., Hintze, M. A., & Rowlett, P. (2003). *The French language today: A linguistic introduction*. London: Routledge.

Behne, D. M., Wang, Y., Alm, M., Arntsen, I., Eg, R., & Valsø, A. (2007). Changes in audio-visual speech perception during adulthood. In *Proceedings of the International Conference on Auditory-Visual Speech Processing* (Vol. 2007). Hilvarenbeek, The Netherlands.

Benguerel, A., & Pichora-Fuller, M. (1982). Coarticulation effects in lipreading. *Journal of Speech, Language and Hearing Research*, *25*(04), 600–607.

Benoît, C., Lallouache, T., Mohamadi, T., Tseva, T., & Abry, C. (1991). Nineteen (±two) French visemes for visual speech synthesis. In *The ESCA Workshop on Speech Synthesis.* (pp. 253–256).

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.

Bergeson, T. R., Houston, D. M., & Miyamato, R. T. (2010). Effects of congenital hearing loss and cochlear implantation on audiovisual speech perception in infants and children. *Restorative Neurology and Neuroscience*, *28*(02), 157–165.

Bergeson, T. R., Pisoni, D. B., & Davis, R. A. O. (2003). A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. *The Volta Review*, *103*(4), 347–370.

Bernstein, L. E., Auer Jr, E. T., Moore, J. K., Ponton, C. W., Don, M., & Singh, M. (2002). Visual speech perception without primary auditory cortex activation. *Neuroreport*, *13*(3), 311–315.

Bernstein, L. E., Tucker, P. E., & Demorest, M. E. (2000). Speech perception without hearing. *Perception & Psychophysics*, *62*(2), 233–252.

Berri, A., & Pagel, D. (2003). Realization of French nasal vowels by Brazilian speakers. *Travaux de l'Institut de Phonetique de Strasbourg*, 105–117.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words*. Cambridge, MA: MIT Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Baltimore, MD: York Press.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*, 305–330.

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.

Binnie, C. A., Montgomery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech and Hearing Research*, *17*, 619–630.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*, 9–49.

Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*(03), 303–328.

Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14*, 131–156.

Bohn, O.-S., & Flege, J. E. (1997). Perception and production of a new vowel category by adult second language learners. In A. James & J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 53–74). Berlin: Walter de Gruyter.

Bohn, O.-S., & Steinlen, A. K. (2003). Consonantal context affects cross-language perception of vowels. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2289–2292.

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*, 447–465.

Borden, G., Gerber, A., & Milsark, G. (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning*, *33*, 499–526.

Bovo, R., Ciorba, A., Prosser, S., & Martini, A. (2009). The McGurk phenomenon in Italian listeners. *Acta Otorhinolaryngologica Italica : Organo Ufficiale Della Società Italiana Di Otorinolaringologia E Chirurgia Cervico-Facciale*, *29*(4), 203–208.

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., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, *112*(1), 272–284.

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.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*(4), 204–220.

Callan, D. E., Callan, A. M., Kroos, C., & Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: A single-sweep EEG case study. *Cognitive Brain Research*, *10*(03), 349–353.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P., … David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593–596.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America*, *54*(2), 421–428.

Carton, F. (1974). *Introduction à la phonetique du francais*. Paris, France: Bordas.

Cathiard, M. A., Schwartz, J.-L., & Abry, C. (2001). Asking a naive question about the McGurk Effect: Why does audio give more percepts with visual than with visual ? In *AVSP 2001 International Conference on Auditory-Visual Speech Processing* (pp. 138–142).

Cienkowski, K. M., & Carney, A. E. (2002). Auditory-visual speech perception and aging. *Ear and Hearing*, *23*(5), 439–449.

Colin, C., Radeau, M., & Deltenre, P. (1998). Intermodal interactions in speech: A French study. In *Proceedings of AVSP'98 International Conference on Auditory-Visual Speech Processing*. Terrigal-Sydney, Australia.

Colin, C., Radeau, M., Deltenre, P., & Morais, J. (2001). Rules of intersensory integration in spatial scene analysis and speechreading. *Psychologica Belgica*, *41*(03), 131–144.

Cummins, P. W. (1981). Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics*, *2*, 131–149.

Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *The Journal of the Acoustical Society of America*, *124*(2), 1264–1268.

Cutler, A., Smits, R., & Cooper, N. (2005). Vowel perception: Effects of non-native language vs. non-native dialect. *Speech Communication*, *47*(1-2), 32–42.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, *116*(6), 3668–3678.

Daly, N., Bench, J., & Chappell, H. (1996). Gender differences in speechreadability. *Journal of the Academy of Rehabilitative Audiology*, *29*, 27–40.

Dancer, J., Krain, M., Thompson, C., & Davis, P. (1994). A cross-sectional investigation of speechreading in adults: Effects of age, gender, practice, and education. *The Volta Review*, *96*, 31–40.

Demorest, M. E., Bernstein, L. E., & DeHaven, G. P. (1996). Generalizability of speechreading performance on nonsense syllables, words, and sentences: subjects with normal hearing. *Journal of Speech, Language and Hearing Research*, *39*, 697–713.

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

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *Tesol Quarterly*, *39*(03), 379–397.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*(2), 121–144.

Diehl, R. L., McCusker, S. B., & Chapman, L. S. (1981). Perceiving vowels in isolation and in consonantal context. *The Journal of the Acoustical Society of America*, *69*(1), 239–248.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In J. Doughty, Catherine & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Oxford, UK: Blackwell.

Durand, J. (1988). Les phénomènes de nasalité en français du Midi: Phonologie de dépendance et sousspécification. *Recherches Linguistiques*, *17*, 29–54.

Erber, N. P. (1971). Effects of distance on the visual reception of speech. *Journal of Speech and Hearing Research*, *14*, 848–857.

Erber, N. P. (1974). Visual perception of speech by deaf children: Recent developments and continuing needs. *Journal of Speech and Hearing Disorders*, *39*(02), 178–185.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Unpublished doctoral dissertation, Utrecht University, Utrecht, the Netherlands.

Feld, J. E., & Sommers, M. S. (2009). Lipreading, processing speed, and working memory in younger and older adults. *Journal of Speech, Language and Hearing Research*, *52*(6), 1555–1565.

Fisher, C. G. (1968). Confusions among visually perceived consonants. *Journal of Speech and Hearing Research*, *11*, 796–804.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65.

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. (1991a). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Heubner & C. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theory* (pp. 249–289). Philadelphia, PA: John Benjamins.

Flege, J. E. (1991b). The interlingual identification of Spanish and English vowels: Orthographic evidence. *The Quarterly Journal of Experimental Psychology*, *43*(3), 701–731.

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. (1997). English vowel production by Dutch talkers: More evidence for the" similar" vs." new" distinction. In A. James & J. Leather (Eds.), *Second-language speech, structure and process.* (pp. 11–52). Berlin: Mouton de Gruyter.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development:*

Papers in honor of Henning Wode (pp. 217–243). Trier, Germany: Wissenschaftlicher Verlag.

Flege, J. E. (2003). Methods for assessing the perception of vowels in a second language. In E. Fava & A. Mioni (Eds.), *Issues in clinical linguistics* (pp. 19–44.). Padova: UniPress.

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., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*, 527–552.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*(1), 1–34.

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.

Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech, 38*(1), 25–55.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78–104.

Florentine, M., Buus, S., Scharf, B., & Canevet, G. (1984). Speech reception thresholds in noise for native and non-native listeners. *The Journal of the Acoustical Society of America, 75*, S84.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*, 3–28.

Fowler, C. A. (1989). Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology, 1*(2), 145–160.

Fowler, C. A., & Rosenblum, L. D. (1990). Duplex perception: A comparison of monosyllables and slamming doors. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 742–754.

Fowler, C. A., & Smith, M. R. (1986). Speech perception as "vector analysis": An approach to the problems of invariance and segmentation. In J. S. Perkell & D. H. Klatts (Eds.), *Invariance and variability in speech processes* (pp. 123–139). Hillsdale, NJ: Erlbaum.

Fox, R. A. (1982). Individual variation in the perception of vowels: Implications for a perception-production link. *Phonetica, 39*, 1–22.

Fuster-Duran, A. (1996). Perception of conflicting audio-visual speech: An examination across Spanish and German. In D. G. Stork & M. E. Hennecke (Eds.), *Speechreading by humans and machines* (pp. 135–143). Berlin: Springer.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(03), 361–377.

Galvao, M. J. C. (1998). The nasal vowels of Iberian Portuguese. *Journal of Acoustical Society of America*, *103*(5), 3087–3087.

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445–2454.

Garcia Lecumberri, M. L., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, *52*(11-12), 864–886.

Gardner, R. C., Masgoret, A. M., Tennant, J., & Mihic, L. (2004). Integrative motivation: Changes during a year-long intermediate-level language course. *Language Learning*, *54*(1), 1–34.

Gass, S. M. (1984). Development of speech perception and speech production abilities in adult second language learners. *Applied Psycholinguistics*, *5*(01), 51–74.

Gelder, B. D., Bertelson, P., Vroomen, J., & Chen, H. C. (1995). Interlanguage differences in the McGurk effects for Dutch and Cantonese listeners. In *Proceedings of the Fourth European Conference on Speech Communication and Technology* (pp. 1699–1702). Madrid, Spain.

Golestani, N., Rosen, S., & Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: The contribution of semantics. *Bilingualism: Language and Cognition*, *12*(3), 1–12.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia*, *9*(3), 317–323.

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics*, *12*, 91–114.

Gottfried, T. L., & Strange, W. (1980). Identification of coarticulated vowels. *The Journal of the Acoustical Society of America*, *68*(6), 1626–1635.

Grant, K. W., & Seitz, P. F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *The Journal of the Acoustical Society of America*, *104*(4), 2438–2450.

Grassegger, H. (1995). McGurk effect in German and Hungarian listeners. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 210–213). Stockholm, Sweden.

Green, K. P., & Gerdman, A. (1995). Cross-modal discrepancies in coarticulation and the integration of speech information: The McGurk effect with mismatched vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(06), 1409–1426.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: female faces and male voices in the McGurk effect. *Perception & Psychophysics*, *50*(6), 524–536.

Hardison, D. M. (1996). Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk Effect. *Language Learning*, *46*, 3–73.

Hardison, D. M. (1998). *Acquisition of second-language speech: Effects of visual cues, context and talker variability*. Unpublished doctoral dissertation, Indiana University, Bloomington, IN.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(04), 495–522.

Hardison, D. M. (2005a). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, *26*(04), 579–596.

Hardison, D. M. (2005b). Variability in bimodal spoken language processing by native and nonnative speakers of English: A closer look at effects of speech style. *Speech Communication*, *46*(1), 73–93.

Hardison, D. M. (2007). The visual element in phonological perception and learning. In M. C. Pennington (Ed.), *Phonology in context* (pp. 135–158). New York, NY: Palgrave Macmillan.

Hayashi, Y., & Sekiyama, K. (1998). Native-foreign langage effect in the McGurk effect: A test with Chinese and Japanese. In *AVSP'98 International Conference on Auditory-Visual Speech Processing.*

Hazan, V., Kim, J., & Chen, Y. (2010). Audiovisual perception in adverse conditions: Language, speaker and listener effects. *Speech Communication*, *52*, 996–1009.

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.

Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, *5*(1), 64–86.

Huensch, A. (2013). *The perception and production of palatal codas by Korean L2 learners of English*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Inceoglu, S. (2011). Perceptual confusability of French nasal vowels. *The Journal of the Acoustical Society of America*, *130*(04), 2573–2573.

Inceoglu, S. (2012). French nasal vowels: Effects of word familiarity on auditory-visual perception. In *Paper presented at the American Association of Applied Linguistics conference*. Boston, MA.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, *126*(2), 866–877.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267–3278.

Iverson, P., Pinet, M., & Evans, B. G. (2011). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(1), 1–16.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics*, *40*(04), 205–215.

Jesse, A., & Massaro, D. W. (2010). Seeing a singer helps comprehension of the song's lyrics. *Psychonomic Bulletin & Review*, *17*(3), 323–328.

Jiang, J., Alwan, A., Keating, P. A., Auer Jr, E. T., & Bernstein, L. E. (2002). On the correlation between facial movements, tongue movements and speech acoustics. *Journal on Applied Signal Processing*, *11*, 1174–1188.

Jiang, J., Chen, M., & Alwan, A. (2006). On the perception of voicing in syllable-initial plosives in noise. *The Journal of the Acoustical Society of America*, *119*(2), 1092–1105.

Johnson, F. M., Hicks, L. H., Goldberg, T., & Myslobodsky, M. S. (1988). Sex differences in lipreading. *Bulletin of the Psychonomic Society*, *26*(2), 106–108.

Johnson, K., Strand, E. a, & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384.

Kellerman, S. (1990). Lip service: The contribution of the visual modality to speech perception and its relevance to the teaching and testing of foreign language listening. *Applied Linguistics*, *11*(03), 239–258.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*(2), 313–334.

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *The Modern Language Journal*, *97*(3), 720–744.

Kleber, F., Harrington, J., & Reubold, U. (2011). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, *55*(3), 383–405.

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.

Lachs, L., Pisoni, D. B., & Kirk, K. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear and Hearing*, *22*(03), 236–251.

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*(02), 227–247.

Lappin, K. (1982). Le français parlé au Québec. *Revue Québécoise de Linguistique*, *11*(2), 93–112.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York, NY: Routledge.

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, *128*(6), 3757–3768.

Lenneberg, E. (1967). *Biological foundations of language*. New York, NY: Wiley & Sons.

Léon, M., & Léon, P. (2007). *La prononciation du français*. Paris: Armand Colin.

Léon, P. (1983). Les voyelles nasales et leurs réalisations dans les parlers français du Canada. *Langue Française*, *60*, 48–64.

Levy, E. S. (2004). *Effects of language experience and consonantal context on perception of French front rounded vowels by adult American English learners of French*. Unpublished doctoral dissertation, The City University of New York, NY.

Levy, E. S. (2009a). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America*, *125*(2), 1138–1152.

Levy, E. S. (2009b). On the assimilation-discrimination relationship in American English adults' French vowel learning. *The Journal of the Acoustical Society of America*, *126*(5), 2670–2682.

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, *36*(1), 141–157.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187–196.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, *35*, 1773–1781.

Lisker, L., & Rossi, M. (1992). Auditory and visual cueing of the [±rounded] feature of vowels. *Language and Speech*, *35*(4), 391–417.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. a, Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076–2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–86.

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore, MD: York Press.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*(03), 251–285.

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

Ludman, C., Summerfield, Q., Hall, D., Elliott, M., Foster, J., Hykin, J. L., … Morris, P. G. (2000). Lip-reading ability and patterns of cortical activation studied using fMRI. *British Journal of Audiology*, *34*, 225–230.

Macchi, M. J. (1980). Identification of vowels spoken in isolation versus vowels spoken in consonantal context. *The Journal of the Acoustical Society of America*, *68*(6), 1636–1642.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, *24*(3), 253–257.

Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception & Psychophysics*, *46*(2), 187–200.

MacLeod, A., & Summerfield, Q. (1990). A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology*, *24*, 29–43.

MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., … Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: An event-related fMRI study. *Neuroreport*, *11*(8), 1729–1733.

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

Manjarrez, E., Mendez, I., Martinez, L., Flores, A., & Mirasso, C. R. (2007). Effects of auditory noise on the psychophysical detection of visual signals: Cross-modal stochastic resonance. *Neuroscience Letters*, *415*(3), 231–236.

Martin, P., Beaudoin-Bégin, A.-M., Goulet, M.-J., & Roy, J.-P. (2001). Les voyelles nasales en français du québec. *La Linguistique*, *37*(2), 49.

Martinet, A. (1988). The internal conditioning of phonological changes. *La Linguistique*, *24*(2), 7–26.

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development*, *55*(5), 1777–1788.

Massaro, D. W. (1987). Speech perception by ear and eye. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip reading* (pp. 53–83). Hillsdale, NJ: Lawrence Erlbaum.

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

Massaro, D. W., & Cohen, M. M. (1993). Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables. *Speech Communication, 13*, 127–134.

Massaro, D. W., Cohen, M. M., Gesi, A., & Heredia, R. (1993). Bimodal speech perception: An examination across languages. *Journal of Phonetics*, *21*, 445–478.

Massaro, D. W., & Light, J. (2003). Read my tongue movements: Bimodal learning to perceive and produce non-native speech /r/ and /l/. In *Proceedings of Eurospeech (Interspeech), 8th European Conference on Speech Communication and Technology*. (pp. 2249–2252). Geneva, Switzerland.

Mateus, M. H., & D'Andrade, E. (2000). *The phonology of Portuguese*. Oxford, UK: Oxford University Press.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language and Hearing Research*, *40*(03), 686.

McGrath, M., & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. *The Journal of the Acoustical Society of America*, *77*(2), 678–685.

McGrath, M., Summerfield, Q., & Brooke, N. M. (1984). Roles of lips and teeth in lipreading vowels. *Proceedings of the Institute of Acoustics*, *6*(4), 401–408.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

Ménard, L., Dupont, S., Baum, S. R., & Aubin, J. (2009). Production and perception of French vowels by congenitally blind adults and sighted adults. *The Journal of the Acoustical Society of America*, *126*(3), 1406–1414.

Mills, A. E. (1983). Acquisition of speech sounds in the visually-handicapped child. In A. E. Mills (Ed.), *Language acquisition in the blind child: Normal and deficient.* (pp. 45–56). San Diego: College Hill.

Mills, A. E. (1987). The development of phonology in the blind child. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip reading* (pp. 145–162). Hillsdale, NJ: Lawrence Erlbaum.

Montagu, J. (2002). L'articulation labiale des voyelles nasales postérieures du français: Comparaison entre locuteurs français et anglo-américains. In *XXIXèmes Journées d'Etudes sur la Parole* (pp. 24–27). Nancy, France.

Montgomery, A. A., & Jackson, P. L. (1983). Physical characteristics of the lips underlying vowel lipreading performance. *Journal of Acoustical Society of America*, *73*(6), 2134–2144.

Montgomery, A. A., Walden, B. E., & Prosek, R. A. (1987). Effects of consonantal context on vowel lipreading. *Journal of Speech and Hearing Research*, *30*, 50–59.

Morosan, D. E., & Jamieson, D. G. (1989). Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word-position or task. *Journal of Speech and Hearing Research*, *32*, 501–511.

Motohashi-Saigo, M., & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology*, *13*(2), 29–47.

Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, *21*, 81–108.

Munro, M. J., Flege, J. E., & Mackay, I. R. A. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, *17*(03), 313–334.

Musacchia, G., Arum, L., & Nicol, T. (2009). Audiovisual deficits in older adults with hearing loss: Biological evidence. *Ear and Hearing*, *30*(5), 505–514.

Nábělek, A. K., & Donahue, A. M. (1984). Perception of consonants in reverberation by native and non-native listeners. *The Journal of the Acoustical Society of America*, *75*, 632–634.

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research*, *71*(1), 4–12.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language and Hearing Research*, *50*(6), 1496–1509.

Nozawa, T., Frieda, E. M., & Wayland, R. (2003). Discriminability and identification of English vowels by native Japanese speakers in different consonantal contexts. *The Journal of the Acoustical Society of America*, *114*, 2364.

Nozawa, T., Wayland, R., & Frieda, E. M. (2003). Effects of consonantal contexts on the perception of English vowels by experienced and inexperienced Japanese learners of English. *The Journal of the Acoustical Society of America*, *113*, 2330.

Öhrström, N., & Traunmüller, H. (2004). Audiovisual perception of Swedish vowels with and without conflicting cues. In *Proceedings of FoNEtIk 2004* (pp. 40–43).

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. In *AVSP 2001 International Conference on Auditory-Visual Speech Processing* (pp. 149–154).

Owens, E., & Blazek, B. (1985). Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech and Hearing Research*, *28*(3), 381–393.

Oxford, R. L., & Anderson, N. J. (1995). A crosscultural view of learning styles. *Language Teaching*, *28*, 201–215.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, *5*(03), 261–283.

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics*, (11), 79–89.

Pereira, Y. I. (2012). Visual cues in the perception of English vowels by L2 learners with Spanish as L1. In *Paper presented at Bilingual & Multilingual Interaction*. Bangor, UK.

Pereira, Y. I. (2013). Perception of English vowels and use of visual cues by learners of English and English native speakers. In *Proceedings of Meetings on Acoustics* (Vol. 19).

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191–215.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *The Journal of the Acoustical Society of America*, *61*, 1352–1361.

Pohl, J. (1983). Quelques caractéristiques de la phonologie du français parlé en Belgique. *Langue Française*, (60), 30–41.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America*, *119*(3), 1684–1696.

Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, *30*, 271–287.

Racine, I., Detey, S., Schwab, S., & Zay, F. (2010). The production of French nasal vowels by advanced Japanese and Spanish learners of French: A corpus-based evaluation study. In *Proceedings of New Sounds 2010 - Sixth International Symposium on the Acquisition of Second Language Speech*. (pp. 367–372).

Rakerd, B., Verbrugge, R. R., & Shankweiler, D. P. (1984). Monitoring for vowels in isolation and in a consonantal context. *The Journal of the Acoustical Society of America*, *76*(1), 27–31.

Rauber, A., Rato, A., Kluge, D. C., & Santos, G. (2012). TP, v. 3.1 [Application software]. http://www.worken.com.br/tp_regfree.php/.

Reid, J. M. (1995). *Learning styles in the ESL/EFL classroom.* Boston, MA: Heinle & Heinle.

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.

Robert-Ribes, J., Schwartz, J.-L., Lallouache, T., & Escudier, P. (1998). Complementarity and synergy in bimodal speech: Auditory, visual, and audio-visual identification of French oral vowels in noise. *The Journal of the Acoustical Society of America*, *103*(6), 3677–3689.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379–410). Timonium, MD: York Press.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics*, *27*(03), 465–485.

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.

Rosenblum, L. D., Johnson, J. A., & Saldaña, H. M. (1996). Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech and Hearing Research*, *39*(06), 1159–1170.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147–1153.

Ruhlen, M. (1973). Nasal Vowels. *Working Papers on Language Universals*, *12*, 1–36.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V, Lu, S. T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, *127*(1), 141–145.

Sams, M., Manninen, P., & Surakka, V. (1998). McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context. *Speech Communication*, *26*(1-2), 75–87.

Sapon, S. M. (1953). An application of psychological theory to pronunciation problems in second language learning. *The Modern Language Journal*, *36*(03), 111–114.

Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 102, pp. 18748–18750).

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. New York, NY: Newbury House.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics*, *59*(1), 73–80.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research*, *47*(3), 277–287.

Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *The Journal of the Acoustical Society of America*, *90*(4), 1797–1805.

Sekiyama, K., & Tohkura, Y. I. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, *21*(04), 427–444.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261.

Shin, D.-J., & Iverson, P. (2013). Training Korean second language speakers on English vowels and prosody. *Proceedings of Meetings on Acoustics*, *19*, 1–4.

Skehan, P. (1991). Individual difference in second language learning. *Studies in Second Language Acquisition*, *13*, 275–298.

Smith, L. C. (2001). L2 acquisition of English liquids: Evidence for production independent from perception. In X. Bonch-Bruevich, W. . Crawford, J. Hellermann, C. Higgins, & H. Nguyen (Eds.), *Selected Proceedings of the 2000 Second Language Research Forum* (pp. 3–22). Somerville, MA: Cascadilla Press.

Sommers, M. S., Spehar, B., & Tye-Murray, N. (2005). The effects of signal-to-noise ratio on auditory-visual integration: Integration and encoding are not independent. *The Journal of the Acoustical Society of America*, *117*, 2574.

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing*, *26*(3), 263–275.

Son, N. Van, Huiskamp, T. M. I., Bosman, A. J., & Smoorenburg, G. F. (1994). Viseme classifications of Dutch consonants and vowels. *Journal of Acoustical Society of America*, *96*, 1341–1355.

Soto-Faraco, S., Navarra, J., Weikum, W. M., Vouloumanos, A., Sebastián-Gallés, N., & Werker, J. F. (2007). Discriminating languages by speech-reading. *Perception & Psychophysics*, *69*(2), 218–31.

Spehar, B., Tye-Murray, N., & Sommers, M. S. (2004). Time-compressed visual speech and age: A first report. *Ear and Hearing*, *25*(6), 565–572.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.

Stevens, G. (1999). Age at immigration and second language proficiency among foreign-born adults. *Language in Society*, *28*(04), 555–578.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatoryacoustic data. In P. B. Denes & E. E. David Jr. (Eds.), *Human communication, a unified view* (pp. 51–66). New York: McGraw-Hill.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3–45.

Stevens, K. N., & House, A. S. (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America*, *27*(3), 484–493.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech, Language and Hearing*, *6*(2), 111–128.

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America*, (55), 653–659.

Straka, G. (1979). *Les sons et les mots. Choix d'études de phonétique et de linguistique*. Paris: Klincksieck.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. a., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Society of America*, *109*(4), 1691–1704.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131–145.

Strange, W., Edman, T. R., & Jenkins, J. J. (1979). Acoustic and phonological factors in vowel identification. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 643–656.

Strange, W., Levy, E. S., & Law, F. F. (2009). Cross-language categorization of French and German vowels by naive American listeners. *The Journal of the Acoustical Society of America*, *126*(3), 1461–1476.

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In M. Zampini & J. Hansen (Eds.), *Phonology and second language acquisition* (pp. 153–192). Cambridge, UK: Cambridge University Press.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. *The Journal of the Acoustical Society of America*, *60*(01), 213–224.

Strelnikov, K., Rouger, J., Demonet, J.-F., Lagleyre, S., Fraysse, B., Deguine, O., & Barone, P. (2013). Visual activity predicts auditory recovery from deafness after adult cochlear implantation. *Brain: A Journal of Neurology*, *136*(12), 3682–3695.

Studdert-Kennedy, M. (1989). Feature fitting: A comment on K. N. Stevens' "On the quantal nature of speech." *Journal of Phonetics*, *17*, 135–143.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, *77*(03), 234–249.

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.

Summerfield, Q. (1991). Visual perception of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 117–137). Hillsdale, NJ: Erlbaum.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions: Biological Sciences*, *335*(1273), 71–78.

Summerfield, Q., MacLeod, A., McGrath, M., & Brooke, M. (1989). Lips, teeth, and the benefits of lipreading. In A. W. Young & H. D. Ellis (Eds.), *Handbook of research on face processing* (pp. 223–233). Amsterdam: Elsevier.

Summerfield, Q., & McGrath, M. (1984). Detection and resolution of audio-visual incompatibility in the perception of vowels. *The Quarterly Journal of Experimental Psychology*, *36*, 51–74.

Takata, Y., & Nábělek, A. K. (1990). English consonant recognition in noise and in reverberation by Japanese and American listeners. *The Journal of the Acoustical Society of America, 88,* 663–666.

Takeuchi, K., & Arai, T. (2009). Strategy for the production of French nasal vowels by Japanese students. In *Proceedings of the Phonetics Teaching and Learning Conference 2007* (pp. 1–4). London, UK.

Traunmüller, H., & Öhrström, N. (2007). Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics, 35*(2), 244–258.

Trofimovich, P., Baker, W., Flege, J. E., & Mack, M. (2003). Second-language sound learning in children and adults: Learning sounds, words, or both? In *Proceedings of the Boston University Conference on Language Development 27.* (pp. 775–786). Sommerville: Cambridge University Press.

Valkenier, B., Duyne, J. Y., Andringa, T. C., & Başkent, D. (2012). Audiovisual perception of congruent and incongruent Dutch front vowels. *Journal of Speech, Language, and Hearing Research, 55,* 1788–1802.

Van Dommelen, W. A., & Hazan, V. (2010). Perception of English consonants in noise by native and Norwegian listeners. *Speech Communication, 52*(11-12), 968–979.

Van Dommelen, W. A., & Hazan, V. (2012). Impact of talker variability on word recognition in non-native listeners. *The Journal of the Acoustical Society of America, 132*(3), 1690–1699.

Violin-Wigent, A. (2006). Southeastern French nasal vowels: Perceptual and acoustic elements. *The Canadian Journal of Linguistics, 51*(1), 15–43.

Walden, B. E., Prosek, R. A., Montgomery, A. A., Scherr, C. K., & Jones, C. J. (1977). Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research, 20*(1), 130–145.

Walker, S., Bruce, V., & O'Malley, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. *Perception & Psychophysics, 57*(8), 1124–1133.

Walter, H. (1977). *La phonologie du français.* Paris, France: Presses Universitaires de France.

Wang, X. (2002). *Training Mandarin and Cantonese speakers to identify English vowel contrasts: Long-term retention and effects on production*. Unpublished doctoral dissertation, Simon Fraser University, Vancouver, Canada.

Wang, X., & Munro, M. J. (1999). The perception of English tense-lax vowel by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. In

Paper presented at the 14th International Congress of Phonetic Sciences. San Francisco, CA.

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., Behne, D. M., & Jiang, H. (2009). Influence of native language phonetic system on audio-visual speech perception. *Journal of Phonetics*, *37*(3), 344–356.

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.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–3658.

Warren, D. W., Dalston, R. M., & Mayo, R. (1993). Aerodynamics of nasalization. In M. K. Huffman & R. A. Krakow (Eds.), *Phonetics and phonology: Nasals, nasalization and the velum* (pp. 119–146). Academic Press.

Werker, J. F., Frost, P. E., & McGurk, H. (1992). La langue et les lèvres: Cross-language influences on bimodal speech perception. *Canadian Journal of Psychology*, *46*(4), 551–568.

Werker, J. F., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63.

Yakel, D. A. (2000). *Effects of time-varying information on vowel identification accuracy in visual speech perception*. Unpublished doctoral dissertation, University of California Riverside.

Yamada, R. A., Strange, W., Magnuson, J. S., Pruitt, J. S., & Clarke III, W. D. (1994). The intelligibility of Japanese speakers' productions of American English /r/, /l/, and /w/, as evaluated by native speakers of American. In *Proceedings of the International Conference of Spoken Language Processing* (pp. 2023–2026). Yokohama, Japan: Acoustical Society of Japan.

Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics*, *52*(4), 376–392.

Yehia, H., Rubin, P., & Vatikiotis-Bateson, E. (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication, 26*, 23–43.

Zampini, M. L., & Green, K. P. (2001). The voicing contrast in English and Spanish: The relationship between perception and production. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 23–48). Malden, MA: Blackwell.

Zerling, J.-P. (1989). The three degrees of labialisation of the French steady-state vowels. In *Proceedings of the European Conference on speech Communication and Technology* (pp. 445–448). Paris, France.

Zerling, J.-P. (1990). *Aspects articulatoires de la labialité vocalique en français. Contribution à la modélisation à partir de labio-photographies, labiofilms et films radiologiques. Étude statique, dynamique et contrastive*. Unpublished doctoral dissertation, University of Strasbourg II, Strasbourg, France.