

THE EFFECT OF VISUAL CUES IN THE PERCEPTION OF NONNATIVE CONTRASTS

DENISE CRISTINA KLUGE*

ABSTRACT

This article reviews and discusses some of the factors that may affect an accurate identification of second language sounds as well as two current models of speech perception: the Speech learning Model and the Native Language Magnet. Then, this article reviews and discusses some studies that investigate the effect of visual cues in the perception of nonnative contrasts by second language learners, specially the ones investigating the perception of Brazilian learners of English (KLUGE, REIS, NOBRE-OLIVEIRA & BETTONI-TECHIO, 2009; KLUGE, 2009). In general, the studies show the importance of visual input in the perception of nonnative contrasts, especially in the perception of the visually distinctive ones.

KEY WORDS: speech perception, visual cues, L2 acquisition.

1 INTRODUCTION

Many studies related to the perception of second language (L2) sounds has discussed the influence of the native language (L1) on accurate percetion of the L2 (FLEGE, 1993, 1995; WODE, 1995; BEST, 1995; IVERSON & KUHL, 1995). Moreover, some models of L2 speech perception have discussed the role of accurate perception on accurate production (FLEGE, 1995; BEST, 1995; ESCUDERO, 2005; BEST & TYLER, 2007). According to some studies (SCHMIDT, 1996; HARNBERGER, 2001; BEST, MCROBERTS & GOODELL, 2001; BEST & TYLER, 2007), it is usually believed that, at least in initial stages of L2 learning, adults are language-specific perceivers and that they perceive L2 segments through the filter of their L1 sound system.

* Universidade Federal do Paraná – Setor Litoral, Matinhos-Paraná, doutorado, professora.
E-mail: klugedenise@hotmail.com

As posited by Flege (1981), L2 sounds may be perceived in terms of those of the L1 by the learner, making this perception different from that of a native speaker. For example, sounds that are separate phonemes in an L2 might be merely allophones of the same phoneme in the L1. Furthermore, Flege states that this may influence the production of L2 sounds by a native speaker of this L1 because of the identical mental representation that this speaker has for the two sounds. Flege (1995) argues that L2 speakers may interpret L2 sounds “through the grid” (WODE, 1978, cited in FLEGE, 1995) of their L1. This fact, “virtually ensures that nonnative speakers will perceive at least some L2 vowels and consonants differently than do native speakers” (FLEGE, 1995, p. 237).

Flege (1995) also posits in his Speech Learning Model (SLM) that the perceived relationship between L1 and L2 categories plays an important role in correctly perceiving or producing L2 sounds. According to one of the hypotheses of this model, L1 and L2 sounds are “related perceptually to one another at a position-sensitive allophonic level” and acquisition of L2 sounds depends on the perceived dissimilarity between L1 and L2 sounds (FLEGE, 1995, p. 239).

Flege’s model is concerned with ultimate achievement in L2 pronunciation, a fact that the model claims to be related to the pattern of speech perception L2 listeners present. In fact, it declares that misperception is the major reason for inaccurate segmental production. However, the model proposes that the possibility to acquire a new sound system is continuously present and can be applied to L2 acquisition.

According to the model, the development of new categories for L2 sounds would be affected by two major variables: age of learning and perceived cross-language phonetic distance. With reference to the last variable, it is hypothesized that “the greater the perceived difference of an L2 sound from the closest L1 sound, the more likely that a separate category will be established for the L2 sound” (FLEGE, 1995, p. 264). It is noteworthy that this hypothesis is closely related to the perception model that will be described next, the Native Language Magnet model proposed by Kuhl and colleagues.

The SLM claims that phonetic category formation might be “blocked by the mechanism of equivalence classification” (FLEGE, 1995, p. 239). Equivalence classification is defined by Flege (1996) as “a basic

cognitive mechanism thought to shape both L1 and L2 speech learning” (p. 13). The mechanism of equivalence classification is important in the process of native-language learning as it enables young children to detect phones produced by different speakers, or in different phonetic contexts, as being part of the same category (FLEGE, 1987). However, Flege hypothesizes that equivalence classification “may lead to foreign accent in older children and adults by preventing them from making effective use of auditorily accessible acoustic differences between phones in L1 and L2” (p. 50).

The concept of equivalence classification determines the categorization of the L2 phones as *identical*, *similar* or *new* in relation to the L1 phones. Wode (1995, p. 323) describes L2 *identical*, *new* and *similar* phones as follows: (a) *identical* phones are dealt by pre-existing categories; (b) *similar* L2 sounds are those that are perceived through the pre-existing categories, and are thus easily and quickly acquired, although they tend to undergo transfer of phonological features of the L1 category, and (c) *new sounds* are those which do not exist in the original sound system of the speaker; and because the perceptual space of this sound is not occupied by any categories, the establishment of the new category tend to be successful, though it might take some time.

According to the SLM, L2 sounds “may be at first identified in terms of a positionally defined allophone of the L1” (FLEGE, 1995, p. 263). However, as L2 learners gain experience, they may become able to discern phonetic differences between L2 sounds and their closest L1 counterparts. In this circumstance, a phonetic category representation may be established for the new L2 sound (FLEGE, 1995). To conclude, the SLM posits that the perceived relationship between categories in L1 and L2 plays an important role in accurately perceiving or producing L2 sounds.

The Native Language Magnet (NLM) model (KUHLE, 1991, 1993; KUHLE, WILLIAMNS, LACERDA, STEVENS & LINDBLOM, 1992; KUHLE & IVERSON, 1995) is based on the key concept that speech categories exhibit certain internal structures that propitiate a sound to be perceived as the best exemplar—or prototype—of a phonetic category. Such internal structures would allow a prototypic member of a phonetic category to be “more quickly encoded” and “more durably remembered” (KUHLE, 1991, p. 93) than any other phonetic exemplar.

The model claims that speech perception is an innate ability and that the establishment of prototypes is a consequence of linguistic experience, which would allow L1 speakers to develop phonetic mental representations of L1 speech sounds since the first days of life. In a series of experimental data, Kuhl and colleagues have demonstrated the ontogeny and phylogeny of prototypes and their role on speech perception, either in L1 or in L2, experiments that will be reviewed below.

Kuhl (1991) aimed at investigating “the nature, function, development, and species specificity of speech prototypes” (p. 94) in L1. In order to answer the four proposed questions, she designed a series of 4 experiments with the participation of three types of listeners: adults, infants and monkeys. Kuhl concludes this study (1991) claiming that the perceptual magnet effect can account for findings in L2 speech perception such as the fact that by the age of 10-12 months children are no longer able to discriminate differences between L2 sounds that they could detect earlier in life. That is, by the age of 6 months infants seem to have already tuned their speech perceptual systems into their L1.

To sum up, the author not only concludes that the magnet effect is inherently human, but also suggests that it is strongly affected by linguistic experience, a hypothesis that would be later confirmed by her own and her coworkers’ studies, (KUHLMAN, 1991; KUHLMAN, WILLIAMNS, LACERDA, STEVENS & LINDBLOM, 1992; KUHLMAN, 1993; KUHLMAN & IVERSON, 1995; KUHLMAN, 2000; KUHLMAN, CONBOY, COFFEY-CORINA, PADDEN, RIVERA-GAXIOLA & NELSON, 2008).

Kuhl and colleagues (2008) offer an expanded version of the NLM based on the most recent studies concerning speech perception and language development. The expanded version of model—or NLM-e— specifies four phases of language perception and language development: (a) a period prior to 6 months of age, infants are language-universal perceivers, being able to discriminate any phonetic units in the world’s languages; (b) a period after 6 months old, infants shift from language-universal to a language-specific mode of speech perception; (c) a period after the enhancement and settlement of the language-specific phonetic perception characterized by progress towards the acquisition of words; and (d) the stage in which the neural commitment related to L1 phonetic space is completely stable, so that any future language learning is probably affected by L1 knowledge.

Another aspect of the NLM, which is particularly relevant to the present article, is that the theory holds that speech representation has a multimodal nature—both auditory and visual information help children to establish their native-language phonetic space. In other words, they argue that “the speech representational system is ‘polymodally mapped’ very early in life” (p. 147), as demonstrated by the classic study of McGurk and McDonald (1976) with both children and adults, which will be further described. The multimodal nature of speech perception has generated further analysis into the role of audio/visual integration in both L1 (e.g., SEKIYAMA & TOHKURA, 1991; MASSARO, COHEN & SMEELE, 1996; ÖHRSTRÖM & TRAUNMÜLLER, 2004, 2007) and L2 (e.g., HAYASHI & SEKIYAMA, 1998; HAZAN et al. 2006), and has been one of the motivations of the present article.

2 VISUAL CUES AND THE PERCEPTION OF NONNATIVE CONTRAST

As stated by Rosenblum (2005), “it is becoming increasingly clear that human speech is a multimodal function, usually apprehended by visual (lipreading) as well as auditory (hearing) means” (p. 51). Summerfield (1992) claims that lipreading is helpful to all sighted people, with normal or impaired hearing, as it compensates rather specifically for the insufficiencies of audition (p. 71). Thus, the speech signal contains multiple acoustic cues to phonetic features and such redundancy of information helps listeners with good hearing in their L1 in many contexts, such as degradation of the signal by noise (SUMBY & POLLACK, 1954); and also help listeners with hearing loss problems (GRANT & SEITZ, 1998a, 1998b).

Concerning blind children, a study conducted by Mills (1987, cited in SCHARTZ, ABRY, BOË & CATHIARD, 2002, p. 264) showed that visual speech input plays an important role on their L1 acquisition. As reported by the authors, research with German, Russian and English blind children has shown that it was difficult for them to learn an “easy-to-see and hard-to hear contrast” such as the nasal consonants /m/ and /n/ in syllable-final position (SCHARTZ et al., 2002, p. 264), for instance.

Grant and Seitz (1998b) define “AV benefit” as the amount of benefit resulting from a combination of auditory and visual cues (p.

2438), and this term has been used to describe the advantage of an audio-visual presentation. However, the AV benefit may depend on the relative perceptual weighting of visual and auditory cues (HAZAN, SENNEMA, FAULKNER, ORTEGA-LLEBARIA, IBA & CHUNG, 2006, p. 1741). One way of evaluating this perceptual weighting is the McGurk effect, which was introduced by McGurk and MacDonald (1976).

In their study, McGurk and MacDonald (1976) investigated audiovisual perception of speech stimuli with conflicting cues. For the stimuli, a woman was filmed while she repeated CV syllables (with interval of 0.5 s) with a stop consonant (/p, b, k, g/) followed by the vowel /a/ (e.g., [papa], [baba], [kaka], [dada]). The audio stimuli were dubbed with different consonants in four combinations as follows: (a) ba-audio /ga-video; (b) ga-audio/ba-video; (c) pa-audio/ka-video; and (d) ka-audio/pa-video. The participants (21 pre-school children, 28 primary school children and 54 adults) were individually tested in two conditions (*Audio/Video* and *Audio only*) and were instructed to repeat what they heard in these conditions.

Results showed that stimuli with conflicting cues such as auditory [baba] and visual [gaga], and auditory [papa] and visual [kaka] were perceived by the participants as [dada] and [tata], respectively, not corresponding to either the auditory or the visual stimulus. Results also showed that some other stimuli, such as auditory [gaga] and visual [baba], was perceived by the participants as [gabga] or [bagba], combining both the auditory and the visual stimuli. These results found by McGurk and MacDonald (1976) suggest that information from auditory and visual modalities are integrated and influence speech perception.

Based on the findings of the McGurk effect, recent studies have investigated the role of visual cues as a variable to investigate the perception of either L1 or L2 contrasts. As for the perception of L1 contrasts, there are some studies such as: (a) Öhström and Traunmüller (2004, 2007), investigating Swedish vowels; (b) Sekiyama & Tohkura (1991), investigating Japanese syllables; and (c) Massaro, Cohen and Smeele, (1996), investigating English CV syllables. Regarding the perception of L2 contrasts, there are some studies such as: (a) Hayashi and Sekiyama (1998), investigating Chinese and Japanese syllables by

Chinese and Japanese speakers; (b) Hazan et al. (2006), investigating English consonants by Spanish and Japanese speakers; (c) Hazan, Sennema and Faulkner (2002) investigating English consonants contrast by Spanish learners of English; and (d) Hardison (1999), investigating English CV syllables by Japanese, Korean, Spanish and Malay speakers. In general, studies have found that both L1 and L2 listeners rely on visual information on the identification of L1 and L2 contrasts, respectively.

A study carried out by Hazan and coworkers (2006) investigated the use of visual cues in the perception of a nonnative consonant contrasts. In their study, Hazan and colleagues investigated the effect of visual salience by evaluating the perception of two English phonemic contrasts differing in the visual distinctiveness of their articulatory gestures: the highly distinctive contrast between labial (/b/-/p/) and labiodental (/v/) consonants (Experiment 1), and the less visually distinctive contrast between /r/ and /l/ (Experiment 2). Both contrasts were tested with Spanish, Japanese and Korean learners of English and also with a group of native-speakers of British English: 32 Spanish, 47 Japanese and 12 English listeners in Experiment 1 and 78 Japanese, 42 Korean and 12 English listeners in Experiment 2.

The participants were tested in three different conditions: a) Video alone (*V*), (b) Audio only (*A*), and (c) Audio-visual (*AV*). Results of Experiment 1 showed that both learner groups achieved higher scores in the *AV* than in the *A* test condition for the highly distinctive contrast, showing evidence of audio-visual benefit. Regarding Experiment 2, results showed that neither group showed evidence of audio-visual benefit for the less visually distinctive contrast. Based on the performance of both learner groups and on the performance of the native speakers of English, Hazan et al. (2006) state that visual salience has an impact on the perception of visual cues to consonant contrasts in both native and nonnative languages, as both native speakers and L2 learners of English achieved much poorer scores in the *V* condition for the less salient /l/-/r/ contrast than the highly salient labial/labiodental contrast, for which near-perfect perception was achieved in the *V* condition for native speakers and even some Spanish-L1 learners of English (p. 1749).

Based on Hazan et al.'s study (2006) as it contains a detailed description of the method used to assess perception, Kluge and colleagues carried out two studies (KLUGE, REIS, NOBRE-OLIVEIRA & BETTONI-TECHIO, 2009; KLUGE, 2009) in order to investigate the use of visual cues in the perception of a visually distinctive English contrast by Brazilian English as Foreign Language (EFL) learners, which will be reviewed in the next section. These were the first studies investigating this variable with Brazilian EFL learners, to the best of my knowledge, as the previous studies have only investigated perception of L2 consonantal sounds auditorily: Koerich, 2002, 2006; Silveira, 2004; Reis, 2006, 2008; Bettoni-Techio, Rauber and Koerich, 2007, Kluge, Rauber, Reis and Bion, 2007; and Moore, 2008, for instance.

3 THE USE OF VISUAL CUES IN THE PERCEPTION OF ENGLISH CONTRASTS BY BRAZILIAN EFL LEARNERS

To investigate the effect of visual cues in the identification of English contrasts by Brazilian EFL learners, the visually distinctive contrast chosen by Kluge and her colleagues in both studies was the word-final nasals /m/ (bilabial) and /n/ (alveolar). In order to understand the difficulties Brazilian learners of English may have with nasal consonants in word-final position, phonological differences between the two languages have to be considered. According to Fujimura and Erickson (1997), typically, nasal consonants have a place distinction between /m/ and /n/ as in English. However, some languages have no place distinction for nasal consonants in the coda (syllable-final position), as Brazilian Portuguese (BP), for instance.

In English, the nasal consonants /m/ and /n/ in word-final position are fully pronounced (O'CONNOR, 1992), with distinct places of articulation (FUJIMURA & ERICKSON, 1997). In fact, these nasal consonants are phonologically distinctive in word-final position, contrasting in minimal pairs such as *Tim-tin*. As explained by O'Connor (1992), the English nasal consonants /m/ and /n/ in word-final position are pronounced by lowering the soft palate and blocking the mouth by closing the lips for /m/; whereas for /n/ the mouth is blocked by pressing the tip of the tongue against the alveolar ridge, and the sides

of the tongue against the sides of the palate. According to O'Connor, the pronunciation of neither of the sounds should cause much difficulty to most speakers. However, she also states that speakers of some languages, such as Portuguese, may have difficulty in pronouncing these nasal consonants in word-final position. O'Connor (1992) explains that "instead of making a firm closure with the lips or tongue tip so that all the breath goes through the nose, they may only lower the soft palate and *not* make a closure, so that some of the breath goes through the nose but the remainder goes through the mouth" (p. 65). When this happens the vowel that precedes the nasal consonant becomes nasalized.

The presence of nasalized vowels or consonants is spread over 99% of the world's languages (CHEN et al., 2007), and this process of coarticulatory nasalization is extremely common. However, the degree of nasalization is different among languages, from subtle as in English (GIEGERICH, 1992; HAMMOND, 1999; LADEFOGED, 2006) to strong as in Portuguese (OLIVEIRA & CRISTÓFARO-SILVA, 2005). It is important to state that although vowel nasalization can occur in English, there are no nasal vowels in this inventory (GIEGERICH, 1992), and nasalization of the vowel does not distinguish the meaning of English words (LADEFOGED, 2005), so nasalization of vowels is not a distinctive feature. Nasalization in BP has provoked different views and theories, but, in general, it is assumed that: (a) phonetically, the nasal consonants /m/ and /n/ are not fully realized after a vowel in word-final position and sometimes not realized at all; and (b) the vowel assimilates nasalization from the following nasal consonant (CRISTÓFARO SILVA, 1999; MATEUS & D'ANDRADE, 2000; CÂMARA JR., 1971; KLUGE ET AL., 2009; KLUGE, 2009). Consideration of the differences in the way the word-final nasal consonants are pronounced in English and BP is very important to the understanding of the difficulties that the Brazilian learners of English may find in the identification of English word-final nasal consonants /m/ and /n/. Moreover, considering these differences between the realization of word-final nasals in English and in BP as well as some of the claims made by the Speech Learning Model which were reviewed in this article, one could expect that Brazilian learners of English would struggle to identify the phonetic dissimilarities between L1 and L2 sounds; and the mechanism of equivalence classification may block accurate perception of the nasals in the L2.

As regards the identification of English word-final nasal by Brazilian EFL learners and in accordance with the other speech model reviewed in this article (Native Language Magnet), one could assume that by around six months old BP speakers would have created their prototypes for the nasals /m/ and /n/ in word-final position according to this language ambient input—through vowel nasalization and nasal deletion—and not fully realized as in English. Bearing in mind that the concept of speech prototypes is the main claim of the NLM, and that the idea of speech prototypes concerns the best instance of certain speech category, one should consider that in the case of BP, the best instance of a word-final /m/ and /n/ is one in which the phonological processes of vowel nasalization and nasal deletion occur. These prototypes would guide the BP listeners' perception, acting as perceptual magnets in the perception of word-final nasals, particularly in the initial stages of L2 learning, a process that might lead to misidentification.

In the first study, Kluge et al. (2009) examined the identification of word-final nasals /m/ and /n/ by ten intermediate Brazilian learners of English assessed by means of a Three-condition Identification Test. In this test, the monosyllabic CVC words with either /m/ or /n/ in word-final position (*Tim/tin*, *gem/gen* and *cam/can*), produced by a male native speaker of English, were presented in three different conditions (a) *Audio only*, in which the participants could only hear the realization of a word; (b) *Audio/Video*, in which the participants could hear and see the realization of a word; and (c) *Video only*, in which the participants could only see the realization of a word. The participants were asked to indicate which of the two nasal consonants they heard and/or saw. There were two blocks of 18 items per condition; giving a total of 108 tokens. Thus, each was repeated three times in each block. Following Hazan et al. (2006), there were two different order of presentation which were counterbalanced: *AV*, *A*, *V* or *A*, *AV*, *V*. According to Hazan and colleagues, the *V only* condition was always presented last because it is likely to be the most difficult condition for the participants. However, Kluge et al.'s results showed that the *V only* condition was not the most difficult one for the identification of the English word-final nasals /m/ and /n/. Results also showed that the *Audio/Video* condition seemed to favor the accurate identification of both word-final nasal consonants

when compared to the *Audio only* condition. Moreover, results showed a slight tendency for the *Audio only* condition to disfavor the accurate identification of both bilabial and alveolar nasal consonants compared to the *Audio/Video* condition.

In general, results indicated that the Brazilian participants seemed to benefit from the *Audio/Video* presentation as discussed by Grant and Seitz (1998b), in the accurate identification of English word-final /m/ and /n/. These results also seem to be in the direction of those of Hazan and colleagues (2006) and suggest that Brazilian learners of English benefited from the *Audio/Video* presentation as the bilabial/alveolar contrast investigated in the pilot study is a visually distinctive contrast.

In this study, Kluge et al. (2009) also investigated the effect of preceding vowel in the identification of English word-final nasals /m/ and /n/ by the Brazilian participants as the literature has shown that phonological context does affect the perception of the target nasals (SHARF & OSTREICHER, 1973; KUROWSKI & BLUMSTEIN, 1984; REPP, 1996; ZEE, 1981, cited in KUROWSKI & BLUMSTEIN, 1995, p. 199; KLUGE, 2004, 2007). However, due to limited number of tokens, they analyzed this variable considering all three conditions tested. Results showed that, among the preceding vowels of the study (/i, ε, æ/), the low previous vowel favored the identification of the English word-final nasal /n/, whereas the high previous vowel disfavored the accurate identification of /m/ in word-final position.

Based on the findings and on the limitations of the previous study, Kluge (2009) further investigate the use of visual cues in the identification of word-final nasal /m/ and /n/ in syllable final position by 42 Brazilian intermediate EFL learners (21 men and 21 women) and by 10 Americans (5men and 5 women), whose data were used as a reference for comparison.

Perception was assessed by means of a Three-condition Identification Test, which contrasted the presence and/or absence of visual cues in the identification of /m/ and /n/ through three types of stimuli presentation—*Audio/Video (AV)*, *Video only (V only)*, and *Audio only (A only)*. The effect of preceding vowels on the identification of the target consonants was controlled through the use of the six words *Tim-tin*, *gem-gen*, and *cam-can* produced by a male native speaker of

English who was video recorded in a soundproof room. The native speaker's mouth was fully visible in the frame during the recording of each item.

The test taken by the participants consisted of 48 items for each of the three conditions (a total of 144 tokens) per participant. This test was conducted individually and the order of the items of each condition was randomized for each participant in order to minimize any ordering effect. Based on the findings of Kluge et al.'s study which showed that *Video only* condition was not the most difficult one for the identification of the English word-final nasals /m/ and /n/, in Kluge (2009)'s study, there were six orders of presentation of the three conditions: (a) *A only, AV, V only*; (b) *A only, V only, AV*; (c) *AV, V only, A only*; (d) *AV, A only, V only*; (e) *V only, A only, AV*; and (f) *V only, AV, A only*. The six orders were counterbalanced; that is, seven participants performed the identification test in each one of the six different orders of presentation. The task of the participants in each trial of each condition was to click on the button corresponding to the English word-final nasal (/m/ or /n/) they heard and/or saw. See Figures 1 and 2 for examples.

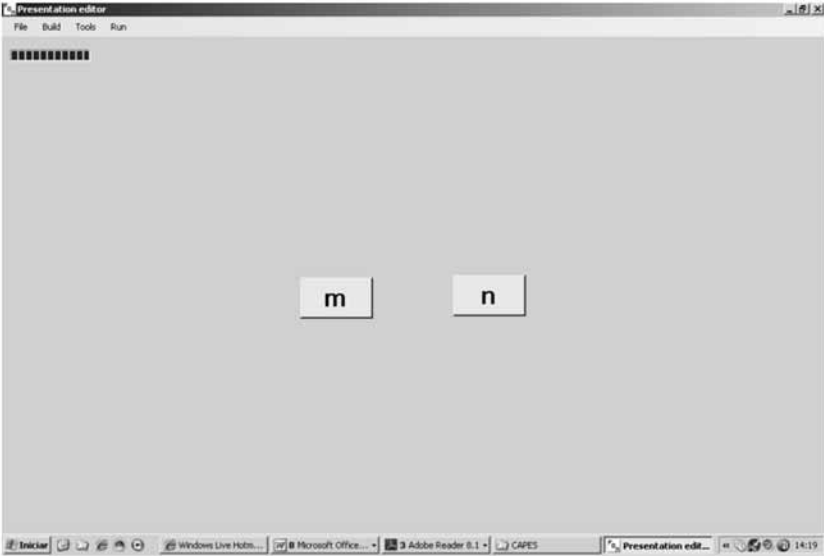


Figure 1. Example of *A only* condition.

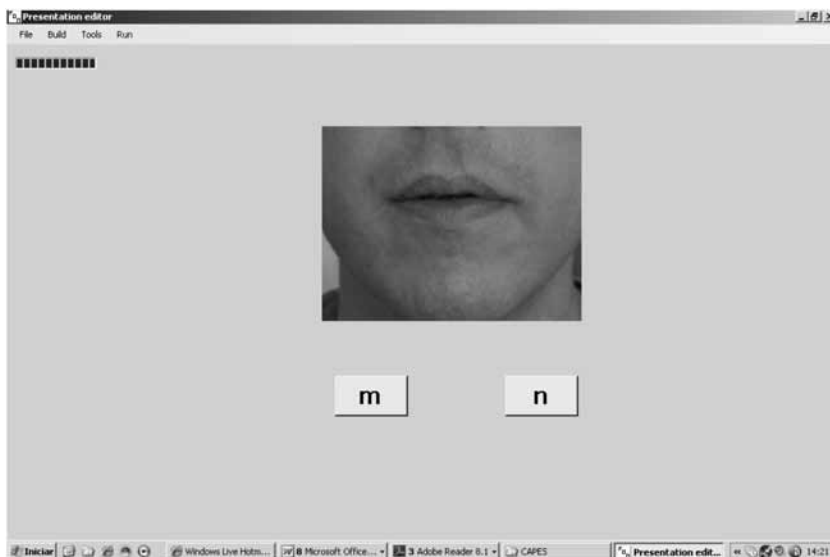


Figure 2. Example of *AV* and *V only* conditions.

Results showed that the Brazilian participants obtained higher scores in two conditions with video input (*Audio Video* and *V only*) indicating that visual cues seemed to favor the accurate identification of English word-final nasals /m/ and /n/ not only in *AV* condition, as showed in Kluge et al. (2009), but also in the *V only* condition. The same tendency was found for the American listeners. Results also revealed that the *A only* condition disfavored the accurate identification of both word-final nasals when compared not only to *AV*, supporting the results of previous study (KLUGE et al., 2009); but also when compared to *V only* condition. The same tendency was found for the American listeners but only in the identification of word-final /n/.

Differently from Kluge et al.'s study (2009) which analyzed the effect of preceding vowel considering all three conditions tested, Kluge (2009) investigated this variable in each of the three conditions. Results revealed that there was an effect for preceding vowel in the identification of /m/ and /n/ only in the *A only* and *AV* conditions for the Brazilian listeners. Whereas the mid preceding vowel favored the accurate identification of /m/, the high preceding vowel disfavored it; these results for the high preceding vowel are in conformance to

those of previous studies (KUROWSKI & BLUMSTEIN, 1995; KLUGE et al., 2009). As for word-final /n/, the mid preceding vowel disfavored the accurate identification of this nasal consonant by the Brazilians in the *A only* and *AV* conditions. These results are contrary to those of Kluge et al. (2009), who found the low preceding vowel to favor the accurate identification of /n/ in the Three-condition Identification Test. As for the American listeners, results revealed no effect of the preceding vowel in the identification of either /m/ or /n/ in any of the conditions of the Three-condition Identification Test.

Summarizing the overall results of the Three-condition Identification Test, Kluge (2009) concludes that it was easier for the Brazilian listeners to identify the place of articulation of the word-final nasal consonants visually than auditorily. She also observes that, in fact, BP speakers do not have difficulty in articulating /m/ or /n/ as they distinctively realize these nasal consonants in word-initial position (e.g., *meta* – “goal” *neta* – “granddaughter”). This suggests that the Brazilian learners may be able to transfer the word-initial distinction present in BP to the word-final distinction in English by observing the visual cues present in the production of these nasal consonants by native speakers.

4 FINAL REMARKS

In this article, some factors that may affect an accurate identification of second language sounds as well two current models of speech perception (Speech Learning Model and Native Language Magnet) were reviewed and discussed. This article also reviewed and discussed some studies that investigate the effect of visual cues in the perception of nonnative contrasts by second language learners and revealed that, in general, L2 learners benefit from an audiovisual presentation, especially in the identification of visually distinctive L2 contrast.

This article also showed that there only two studies, to the best of my knowledge, investigating the use of visual cues in the perception of English contrast by Brazilian learners of English. Studies considering Brazilian/English interphonology are very important to contribute to the improvement of pronunciation teaching and the development of

pronunciation materials concerning the BP speakers' specific difficulties concerning English learning.

O EFEITO DAS PISTAS VISUAL NA PERCEÇÃO DE CONTRASTES NÃO NATIVOS

RESUMO

Este artigo revisa e discute alguns fatores que podem afetar a percepção de sons da língua estrangeira, bem como dois modelos de percepção de fala: Modelo de Aprendizagem de Fala (*Speech Learning Model*) e Modelo de Magneto da Língua Materna (*Native Language Magnet*). Este artigo também revisa e discute alguns estudos que investigam o efeito das pistas visuais na percepção de sons da língua estrangeira, especialmente os que investigam a percepção de brasileiros aprendizes de inglês (KLUGE, REIS, NOBRE-OLIVEIRA & BETTONI-TECHIO, 2009; KLUGE, 2009). Em geral, os estudos mostram a importância das pistas visuais na percepção de contrastes não nativos, especialmente na percepção de contrastes visualmente distintos.

PALAVRAS-CHAVE: percepção de fala, pistas visuais, aquisição de L2.

REFERENCES

- BEST, C. T. A direct realistic view of cross-language speech perception. In: STRANGE, W. (Ed.). *Speech perception and linguistic experience: issues in cross-language research*. Timonium, MD: York Press, 1995. p. 171-206.
- BEST, C. T. & TYLER, M. D. Nonnative and second-language speech perception: Commonalities and complementarities. In: BOHN, O. S.; MUNRO, M. J. (Eds.). *Second language speech learning: the role of language experience in speech perception and production*. Amsterdam: John Benjamins, 2007. p. 13-34.
- BEST, C. T.; MCROBERTS, G. W. & GOODELL, E. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, v. 109, n. 2, p. 775-794, 2001.
- CHEN, N. F.; SLIFKA, J. & STEVENS, K. N. Vowel nasalization in American English: acoustic variability due to phonetic context. In: *Proceedings of 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany, 6-10 August, 2007. p. 905-908.
- CÂMARA JR., J. M. *Problemas da lingüística descritiva*. Petrópolis: Vozes, 1971.
- CRISTÓFARO SILVA, T. *Fonética e fonologia do português: roteiro de estudos e guia de exercícios*. São Paulo: Contexto, 1999.

ESCUADERO, P. Linguistic Perception and Second Language Acquisition. Explaining the attainment of optimal phonological categorization. Doctoral dissertation, Utrecht University, 2005.

FLEGE, J. E. The phonological basis of foreign accent: a hypothesis. *TESOL Quarterly*, v. 15, p. 443-455, 1981.

FLEGE, J. E. The production of the “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, v. 15, p. 47-65, 1987.

FLEGE, J. E. Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, v. 93, n. 3, p. 1589-1608, 1993.

FLEGE, J. E. Second language speech learning: theory, findings, and problems. In: STRANGE, W. (Ed.), *Speech perception and linguistic experience: issues in cross-language research*. Timonium, MD: York Press, 1995. p. 233-272.

FLEGE, J. E. English vowel production by Dutch talkers: more evidence for the “similar” vs. “new” distinction. In: JAMES, A.; LEATHER, J. (Eds.). *Second-language speech: structure and process*. Berlin: Mouton de Gruyter, 1996. p. 11-52.

FUJIMURA, O. & ERICKSON, D. Acoustic Phonetics. In: HARDCASTLE, W. J. & LAVER, J. (Eds.). *The handbook of phonetic sciences*. Cambridge: Blackwell Publishers, 1997. p. 65-115.

GIEGERICH, H. J. *English phonology: an introduction*. Cambridge: Cambridge University Press, 1992.

GRANT, K. W. & SEITZ, P. F. Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America*, 103 (5), 2677-2690, 1998a.

GRANT, K. W. & SEITZ, P. F. Measures of auditory-visual integration in nonsense syllables and sentences, *Journal of the Acoustical Society of America*, 104, 2438-2450, 1998b.

HARDISON, D. Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk effect, *Language Learning*, 49, 213-283, 1999.

HARNSBERGER, J. D. On the relationship between identification and discrimination of non-native nasal consonants. *Journal of the Acoustical Society of America*, 110, 489-503, 2001.

- HAZAN, V.; SENNEMA, A.; FAULKNER, A. Audiovisual perception in L2 learners. *Proceedings of the International Conference for Spoken Language Processing*. Denver, 2002, p. 1685-1688.
- HAZAN, V.; SENNEMA, A.; FAULKNER, A.; ORTEGA-LLEBARIAD, M.; IBA, M.; CHUNG, H. The use of visual cues in the perception of non-native consonant contrasts. *Journal Acoustical Society of America*, v. 119, n. 3, p. 1740-1751, 2006.
- IVERSON, P. & KUHL, P. K. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, v. 97, p. 533-562, 1995.
- KLUGE, D. C. *The perception and production of English syllable-final nasals by Brazilian learners*. Unpublished master thesis. Florianópolis: Universidade Federal de Santa Catarina, 2004.
- KLUGE, D. C. *Brazilians' perception and production of English word-final nasals*. Unpublished paper. Florianópolis: Universidade Federal de Santa Catarina, 2007.
- KLUGE, D. C.; RAUBER, A. S.; REIS, M. S.; BION, R. A. H. The relationship between perception and production of English nasal codas by Brazilian learners of English. *Proceedings of Interspeech*, 2007. p. 2297-2300.
- KLUGE, D. C.; REIS, M. S.; NOBRE-OLIVEIRA, D. & RAUBER, A. S. Intelligibility of accented speech: the perception of word-final nasals by Dutch and Brazilians. *V Jornadas en Tecnología del Habla*, Bilbao, 2008. p. 199-202.
- KLUGE, D. C.; REIS, M. S.; NOBRE-OLIVEIRA, D.; BETTONI-TECHIO, M. The use of visual cues in the perception of English syllable-final nasals by Brazilian EFL learners. In: WATKINS, M. A.; RAUBER, A. S. & BAPTISTA, B. O. (Eds.). *Recent Research in Second Language Phonetics/Phonology: Perception and Production*. Cambridge Scholars Publishing, 2009. p. 141-153.
- KLUGE, D. C. Brazilian EFL learners' identification of word-final /m-n/: native/nonnative realizations and effect of visual cues. Unpublished doctoral dissertation. Universidade Federal de Santa Catarina: Florianópolis, 2009.
- KOERICH, R. D. Perception and production of word-final vowel epenthesis by Brazilian EFL students. Unpublished doctoral dissertation. Universidade Federal de Santa Catarina: Florianópolis, 2002.
- KOERICH, R. D. Perception and production of vowel paragoge by Brazilian EFL students. In: BAPTISTA, B. O. & WATKINS, M. A. (Eds.). *English with a Latin Beat: Studies in Portuguese/Spanish – English Interphonology*. Amsterdam: John Benjamins, 2006. p. 91-104.

- KUHL, P. K. Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics*, v. 21, p. 125-139, 1993.
- KUHL, P. K. A new view of language acquisition. *Proceedings of the National Academy of Sciences*, v. 97, p. 11850-11857, 2000.
- KUHL, P. K. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, v. 50, p. 93-107, 1991
- KUHL, P. B.; CONBOY, B. T.; COFFEY-CORINA, S.; PADDEN, D.; RIVERA-GAXIOLA, M. & NELSON, T. Phonetic learning as a pathway to language: new data and native language model theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*, v. 363, p. 979-1000, 2008.
- KUHL, P. K.; IVERSON, P. Linguistic experience and the “perceptual magnet effect”. In: STRANGE, W. (Ed.). *Speech perception and linguistic experience: issues in cross-language research*. Timonium, MD: York Press, 1995. p. 121-154.
- KUHL, P. K., WILLIAMS, K. A., LACERDA, F., STEVENS, K. N. & LINDBLOM, B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, v. 255, p. 606-608, 1992.
- KUROWSKI, K. M. & BLUMSTEIN, S. E. Acoustic properties for place of articulation in nasal consonants. *Journal of the Acoustical Society of America*, v. 81, n. 6, p. 1917-1927, 1987.
- KUROWSKI, K. M. & BLUMSTEIN, S. E. Acoustic properties for the perception of nasal consonants. In: HUFFMAN, M. K.; KRAKOW, R. A. (Eds.). *Nasals, Nasalization and the Velum* Academic Press, Inc., 1995. p. 197-222.
- LADEFOGED, P. *Vowels and consonants: an introduction to the sounds of Languages*. Second Edition. Oxford: Blackwell Publishing, 2005.
- LADEFOGED, P. *A course in phonetics*. Fifth Edition. USA: Thomson Wadsworth, 2006.
- MASSARO, D., COHEN, M. M. & SMEELE, P. M. T. Perception of asynchronous and conflicting visual and auditory speech. *Journal of the Acoustical Society of America*, v. 100, n. 3, p. 1777-1786, 1996.
- MATEUS, M. H. M., & D’ANDRADE, E. *The phonology of Portuguese*. Oxford: Oxford University Press. 2000. 376p.
- MCGURK, H. AND MACDONALS, J. Hearing lips and seeing voices. *Nature*, v. 264, p. 746-748, 1976.

- MOORE JR., D. H. The perception of English word-final /l/ by Brazilian learners. Unpublished master's thesis. Florianópolis: Universidade Federal de Santa Catarina, 2008.
- O'CONNOR, J. D. *Better English pronunciation*. Cambridge: Cambridge University Press, 1992.
- ÖHRSTRÖM, N., TRAUNMÜLLER, H. Audiovisual perception of Swedish vowels with and without conflicting cues. *Proceedings, FONETIK 2004*, Department of Linguistics, Stockholm University, 2004.
- ÖHRSTRÖM, N., TRAUNMÜLLER, H. Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics*, v. 35, p. 244-258, 2007.
- OLIVEIRA, J. C. & CRISTÓFARO-SILVA, T. Aprendizado de língua estrangeira: o caso da nasalização de vogais. Unpublished paper. Universidade Federal de Minas Gerais. Available at <<http://www.cori.unicamp.br/jornadas/completos/UFGM/ND1010.doc>>, 2005.
- REIS, M.S. The perception and production of English interdental fricatives by Brazilian EFL learners. Unpublished master's thesis, Universidade Federal de Santa Catarina, 2006.
- REIS, M.S. The perception of English voiceless interdental by four language backgrounds. Unpublished paper, Universidade Federal de Santa Catarina, 2008.
- REIS, M. S. & KLUGE, D.C. Intelligibility of Brazilian Portuguese-Accented English Realization of Nasals in Word-Final Position by Brazilian and Dutch EFL Learners. *Revista Crop*, v. 13, p. 215-229, 2008.
- ROSENBLUM, L. D. Primacy of multimodal Speech Perception. In: Pisoni, D. B.; Remez, E. R. (Eds.) *The handbook of speech perception*. Blackwell Publishing. p. 51-78, 2005.
- SCHARTZ, J. L.; ABRY, C.; BOË, L. J. & CATHIARD, M. Phonology in a Theory of Perception-for-Action-Control. In: DURAND, J. & LARKS, B. *Phonetics, Phonology, and Cognition*. Oxford: Oxford University Press, 2002. p. 255-280.
- SCHMIDT, A. M. Cross-language identification of consonants. Part 1. Korean perception of English. *Journal of the Acoustic Society of America*, v. 99, n. 5, p. 3301-3211, 1996.
- SHARF, D. J. & OSTREICHER, H. Effects of forward and backward coarticulation on the identification of speech sounds. *Language and Speech*, v. 16, p. 196-206, 1973.

SILVEIRA, R. *The influence of pronunciation instruction on the perception and production of English word-final consonants*. Unpublished doctoral dissertation, Universidade Federal de Santa Catarina, Florianópolis, 2004.

SUMBY, W. H.; POLLACK, I. Visual contribution to speech intelligibility in noise, *Journal of the Acoustical Society of America*, v. 26, p. 212–215, 1954.

SUMMERFIELD, Q. Lipreading and audio-visual speech perception. *Philosophical Transactions: Biological Sciences*, v. 335, n. 1273, p. 71-78, 1992.

WERKER, J. F., & POLKA, L. Developmental changes in speech perception: new challenges and new directions. *Journal of Phonetics*, v. 21, p. 83-101, 1993.

WODE, H. Speech perception, language acquisition and linguistic: Some mutual implications. In: STRANGE, W. (Ed.). *Speech perception and linguistic experience: issues in cross-language research*. Timonium, MD: York Press, 1995. p. 321-350.