The Use of Visual Cues in the Perception of English Syllable-Final Nasals by Brazilian EFL Learners

Denise Cristina Kluge[1], *Mara Silvia Reis*[1], *Denize Nobre-Oliveira*[2], *Melissa Bettoni-Techio*[1]

[1]Federal University of Santa Catarina, Brazil
[2]Federal Center of Technological Education of Santa Catarina, Brazil
klugedenise@hotmail.com, marasreis@hotmail.com,
denizenobre@yahoo.com.br, melissa@formatto.com.br

1. Introduction

As stated by Rosenblum (2005:51), "it is becoming increasingly clear that human speech is a multimodal function, usually apprehended by visual (lipreading) as well as auditory (hearing) means". 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 and Pollack, 1954); and also may help listeners with hearing loss problems (Grant & Seitz, 1998). As defined by Grant and Seitz (1998:2438), "AV benefit" is the amount of benefit resulting from a combination of auditory and visual cues, 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, 2006:1741).

One way of evaluating this perceptual weighting is the McGurk effect (McGurk and McDonald, 1976). The McGurk effect occurs when visual and acoustic cues diverge, for instance, visual /g/ with auditory /b/ was perceived as /d/ by the listeners. As reported by Hazan, Sennema, Iba and Faulkner (2005:362), this resulting perception indicates that "the information from each modality is integrated before the sound is categorized by the listener".

The research carried out by Hazan et al. (2006) on the use of visual cues in the perception of a non-native contrast is particularly relevant to the present study because it deals with perception of consonants and contains a detailed description of the method used to assess perception. The study 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 and Japanese learners of English and also with a group of native-speakers of English. The participants were tested in three different conditions: a) Video alone (*V*), (b) Audio alone (*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, thus showing evidence of audio-visual benefit. As regards 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. (2006:1749) state that visual salience has an impact on the perception of visual cues to consonant contrasts in both native and non-native 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.

Based on previous studies, in this study it is expected that the Brazilian learner of English will benefit from AV presentation as the labial/alveolar contrast in syllable-final position is a visually distinctive contrast in English. Brazilian learners of English have difficulties with English nasals in syllable-final position due to phonological differences between the two languages.

Usually, nasal consonants have a place distinction between / m/ and /n/ as in English (Fujimura & Erickson, 1997). However, some languages have no place distinction for nasal consonants in the coda (syllable-final position), as Brazilian Portuguese (BP), for instance. In BP, it is accepted by most authors, that: (1) phonetically, the nasal consonants /m/ and /n/ are not fully realized after a vowel in syllable-final position and sometimes not realized at all; and (2) the vowel receives nasalization from the following nasal consonant (Cristófaro Silva, 1999; Mateus & d'Andrade, 2000; Câmara Jr., 1971). This is an assumption of the present study. Due to the difference in the pronunciation of the nasal consonants / m/ and /n/ in syllable-final position in English and BP, Brazilian EFL learners may have difficulty in the perception of the English nasals in coda position.

Flege (1995) states in his Speech Learning Model (SLM) that the perceived relationship between L1 and L2 categories plays an important role in accurately perceiving or producing L2 sounds. Flege's (1995:239) model hypothesizes that 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. Therefore, Brazilian learners of English would be expected to have difficulty distinguishing/identifying English /m/ and /n/ in syllable-final position as, in BP, these coda nasal consonants are not distinctive due to vowel nasalization and deletion of the nasal consonant.

Kluge et al. (2007), investigated the effect of vowel context in the perception of English nasal consonants /m/ and /n/ in syllable final position by twenty pre-intermediate Brazilian learners of English and three native speakers of American English as a reference for comparison. Perception was assessed through a Categorial Discrimination Test (Flege, Munro & Fox, 1994) and an Identification Test with only audio conditions. The results revealed that both the perception tests were quite difficult for the Brazilian learners, as students failed to accurately perceive the English coda nasals in less than half of the trials. Results also showed that the previous vowel seemed to influence the accurate perception of the target nasals by both the Brazilian learners and the native speakers. Both groups seemed to have difficulties in either discriminating or identifying the target nasals in the context of nearly the same previous vowels, although to different degrees. The results showed that high vowels tended to disfavor the accurate discrimination of the nasal consonants, while low vowels seemed to favor the accurate perception of English coda nasals.

As reviewed above, research on the perception of nasal consonants in syllable-final position by Brazilian EFL learners has been carried out, but there is no investigation concerned with the effect of visual cues in the identification of the nasal consonants, to be best of our knowledge. Therefore, in order to investigate the perception of the nasal consonants in the coda by Brazilian EFL learners, the main objectives of this study are: (i) to investigate which condition (Audio only, Audio/Video or Video only) favors the accurate identification of the syllable-final nasal /m/ and /n/; and (ii) to investigate whether the phonological context influences the perception of the English syllable-final nasals, considering the previous vowel as a variable.

2. Method

2.1 Participants

Ten participants were tested (7 women and 3 men), ranging in age from 18 to 25. All participants were considered intermediate students from Universidade Federal de Santa Catarina regularly attending either the third semester of the English undergraduate course or the fourth or fifth semester of English in the Extracurricular Language Program. In order to ensure the level of the participants, the students recorded a short free speech about themselves. Two English speakers (one native and the researcher, who is non-native) rated the pronunciation of the participants holistically from non-native-like to close to native-like on a 1-5 scale, 1 being non-native like and 5 close to native-like. If the participants were rated 1 or 5 they would be excluded in order to avoid having low or high proficient participants in English. Thus, the ten participants of this study were rated from 2 to 4. According to the participants' report, everyone in the group resided in Florianópolis, Santa Catarina at the time of data collection. None of them had been to any English speaking country and none had had extensive exposure to English. All the participants also reported that they considered they had had more exposure to American English than to British English, for instance. Six of the ten participants reported that they had never had formal instruction on English sounds and pronunciation and/or on phonetic symbols. All the participants also reported they have no hearing problems.

2.2 Material

The data for the present study was collected in December, 2006 at the Universidade Federal de Santa Catarina (UFSC). For the purpose of data collection, three instruments were designed: two questionnaires (one to assess participants' biographical information, and the other to assess the participants' impressions and opinions about each of the three conditions of the Identification Test concerning degree of difficulty and length of the test); and an Identification test. The test followed the design of the three-condition Identification test described by Hazan et al. (2006) to assess second-language learners' sensitivity to phonetic information contained in visual cues when they have to identify a non-native phonemic contrast.

2.2.1 Three-condition identification test's stimuli

The test contained two blocks of 18 monosyllabic CVC words. The target words were three minimal pairs contrasting /m/ and /n/ in syllable-final position preceded by: /ɪ/ (high), /ɛ/ (mid) or /æ/ (low): *Tim/tin, gem/gen* and *cam/can*.

For the construction of the test, a phonetically trained male speaker of America English recorded the test items. Video recordings were made in a soundproof room on a Canon Elura 40-MC digital camcorder. The talker's mouth was fully visible in the frame during the recording of each item. The camera was set to record in stereo at 44.1 kHz, 16 bits. The video was recorded at full digital video resolution, 720 x 480 pixels. After recording, the video with sound was imported into an Apple Macintosh PowerBook G4 using Apple's consumer video editing program iMovie. In iMovie, the video stream was edited to isolate and separate the individual items into separate video clips, after which the iMovie project was saved and duplicated twice for a total of three copies. The video clips were edited so that the start and end of the pronunciation of each word showed a neutral face expression.

The test items were the same in each of the three conditions, but in the Audio only or in the Video only conditions, either the auditory or visual cues were removed. For the Audio only, one copy of the project was opened, and the audio was extracted from the video and exported into a separate file at its original 16 bit, 44.1 kHz stereo resolution. The audio only file of all words was edited into separate files with Bias software's Peak DV program, and ultimately all words were saved into their own AIFF file. These files were later converted to MP3 with Apple’s iTunes software. For the Video only, the second copy of the project was opened in iMovie, the audio was removed, and the words were exported individually as320 x 240 pixel AVI files. For the Audio and Video, the third copy of the project was opened in iMovie, and the video clip of each word was exported as a 320 x 240 pixel AVI file, with audio at full 16 bit, 44.1 kHz stereo resolution. Thus, the six monosyllabic words with /m/ or /n/ in the coda were presented in three conditions: (a) Audio only (A), in which the participants heard one pronunciation of a word; (b) Audio/Video (AV), in which the participants heard and saw one pronunciation of a word; and (c) Video Alone (V), in which the participants only saw one pronunciation of a word.

In order to verify if the stimuli recorded were good exemplars of the target phonemes, the stimuli from each of the three conditions were randomized and played to an American native speaker of English. Thus, the stimuli used in each condition of this test were those correctly identified by the native speaker of English. There were two blocks of 18 items per condition; giving a total of 108 tokens. Thus, each was repeated three times in each block.

The test was conducted with a Power Point presentation and the order of the items was randomized within each block to minimize any ordering effect. Following Hazan et al. (2006), two orders for the presentation of the three conditions were used: A, AV, V or AV, A, V, and the two orders were counterbalanced. According to the authors, the Video only condition was always presented last because it is likely to be the most difficult condition for the participants.

The task of the participants in each trial of each block of each condition was to circle on an answer sheet the English coda nasal they heard and/or saw. A familiarization task of eight items with different words for each of the three conditions was also designed.

2.2.2 Procedures

Data was collected individually on a laptop computer. First, the participants answered the questionnaire with their biographical information. The data collection took approximately 20 minutes. As there were two orders for presentation, half of the students followed the order: (1) Audio only condition, (2) Questionnaire about the Audio test condition, (3)Audio/Video condition, (4) Questionnaire about the AV test condition, (5) Video alone, (6) Questionnaire about the Video test condition; and the other half followed the order: (1) Audio/Video condition, (2) Questionnaire about the AV test condition, (3)Audio only condition, (4) Questionnaire about the Audio test condition, (5) Video alone, (6) Questionnaire about the Video test condition;

3. Results and discussion

As regards the first objective, Table 1 shows the number of correct identifications and percentages of the target nasal for each of the three conditions by each of the participants. Percentages were based on the total of 18 answers for each of the target nasals in each of the three conditions for each of the participants, resulting in a total of 180 answers for each condition.

Table 1. Number of correct identifications of the target nasals in the Identification test by participant (from a total of 18 tokens of each consonant per test condition)

Partic.	A /m/	AV /m/	V /m/	A /n/	AV /n/	V /n/
1	18(100%)	18(100%)	18(100%)	17(94.44%)	18(100%)	18(100%)
2	12(66.67%)	18(100%)	18(100%)	13(72.22%)	18(100%)	17(94.44%)
3	15(83.33%)	17(94.44%)	18(100%)	17(94.44%)	18(100%)	17(94.44%)
4	18 (100%)	16(88.89%)	13(72.22%)	8(44.44%)	16(88.89%)	11(61.11%)
5	16(88.89%)	17(94.44%)	18(100%)	16(88.89%)	18(100%)	18(100%)
6	17(94.44%)	18(100%)	18(100%)	15(83.33%)	18(100%)	18(100%)
7	14(77.78%)	17(94.44%)	16(88.89%)	12(66.67%)	17(94.44%)	16(88.89%)
8	12(66.67%)	17(94.44%)	14(77.78%)	13(72.22%)	17(94.44%)	15(83.33%)
9	13(72.22%)	15(83.33%)	15(83.33%)	14(77.78%)	17(94.44%)	10(55.55%)
10	12(66.67%)	17(94.44%)	18(100%)	15(83.33%)	18(100%)	14(77.78%)
Total	147 (81.67%)	170 (94.44%)	166 (92.22%)	140 (77.77%)	175 (97.22%)	154 (85.55%)

Taking into consideration the performance of the participants as a group in identifying both coda nasals /m/ and /n/, the results show that the percentage of accurate identification of both target nasal consonants by the participants gradually increased from Audio only to Video only to Audio/Video condition. Differently from the results of Hazan et al. (2006) who found the Video only condition to be the most difficult condition, the results of this study indicate that the Video alone condition did yield a certain degree of difficulty, as the participants accurately identified the target nasal consonant in 320 tokens out of 360, but it was not the most difficult condition.

The Friedman statistical test showed a significant effect for test condition in the identification of /m/ (X^2 (2, N=10) = 8.909, p = .012). Thus, Wilcoxon tests were run in order to verify which pairs of test conditions were significant for the identification of the bilabial nasal in syllable-final position and yielded the following results: (a) for the pair 'Audio only' vs. 'Video only' the difference was not significant (Z = -.1.854, p= .064); (b) for the pair 'Audio/Video' vs. 'Video only' the difference was also not significant (Z = - .649, p= .516); and (c) for the pair 'Audio/Video' vs. 'Audio Only' the difference was significant (Z = -2.203, p= .028). The results indicate that the Audio/Visual condition favored the accurate identification of the English syllable-final nasal /m/, thus corroborating Hazan et al. (2006), and showing the importance of visual input for the perception of a visually distinctive contrast.

The Friedman statistical test also showed a significant effect for test condition in the identification of /n/ (X^2 (2, N=10) = 14.389, p = .001). Therefore, Wilcoxon tests were run and revealed the following results: (a) for the pair 'Audio only' vs. 'Video only' the difference was not significant (Z = -.1.550, p= .121); (b) for the pair 'Audio/Video' vs. 'Video only' the difference was also not significant (Z = - 2.384, p= .017); and (c) for the pair 'Audio/Video' vs. 'Audio Only' the difference was significant (Z = -2.814, p= .005). The results indicate that the Audio/Visual condition favors the accurate identification of the English syllable-final nasal /n/, just as the bilabial nasal, again corroborating Hazan et al. (2006), as hypothesized, and demonstrating the importance of visual input for the perception of a visually distinctive contrast.

These results are also in accordance with the participants' impressions (assessed by means of a questionnaire as explained in 2.2) as regards the difficulty they felt in identifying both nasals in each of the three conditions. Most participants (6 out of 10) reported that the Audio/Video was the easiest condition for the identification of the target nasal because, as they reported, they could see the mouth's movement which helped them to identify the consonant that they heard. As regards the most difficult condition, most participants (8) reported the Audio only to be the most difficult one. One of the reasons reported for the Audio only being the most difficult condition was the fact that the words were not contextualized and they had no other source to help them to decide which nasal consonant they heard. When asked if they pay attention to the movements of the mouth/lips when talking to another person in either BP or English, 6 participants reported that they do in BP when in a noisy situation or when they want to pay better attention to what is been said, and 7 participants reported that they do when speaking English in face to face conversation when they do not understand a specific word or what is being said in general.

As regards the second objective, Table 2 shows the number and percentages of correct identification of the English nasal consonants /m/ and /n/ in syllable-final position in the context of three different previous vowels: high (/ɪ/), mid (/ɛ/) and low (/æ/). The results are based on the analysis of the three conditions of the Identification test.

Table 2. Results of the perception of English syllable-final nasal /m/ and /n/ as regards previous vowel in the Identification test

Previous vowel	No. Answers /m/	No. Correct Answers /m/	No. Answers /n/	No. Correct Answers /n/
/ɪ/	180	147 (81.67%)	180	154 (85.56%)
/ɛ/	180	169 (93.89%)	180	148 (82.22%)
/æ/	180	165 (91.67%)	180	170 (94.44%)
Total	540	481 (89.07%)	540	472 (87.44%)

Table 2 shows that the English coda nasal /m/ was less frequently identified in the context of the high vowel /ɪ/. The results also show that the participants had a similar performance in the context of the mid and the low vowels. The Friedman statistical test showed a significant effect for previous vowel followed by /m/ (X^2 (2, N=10) = 9.267, p = .01). Therefore, Wilcoxon tests were run in order to verify whether the differences between the pairs of previous vowels were significant for the identification of the bilabial nasal in coda position and yielded the following results: (a) for the pair 'low vowel' vs. 'mid vowel' the difference was not significant (Z = -.755, p = .450); (b) for the pair 'high vowel' vs. 'mid vowel' the difference was significant (Z = - 2.536, p = .011); and (c) for the pair 'high vowel' vs. 'low vowel' the difference was also significant (Z = -2.047, p= .041). Thus, the results indicate that there is an effect of the previous high vowel to disfavor the accurate identification of the English coda nasal /m/. These results corroborate those found by Kurowski & Blumstein (1987, 1995) and Kluge et al. (2007).

As regards the identification of the English coda nasal /n/, Table 2 shows that the alveolar coda nasal was more accurately identified in the context of the low vowel /æ/. The results also show that the Brazilian learners had a similar performance in the context of the mid and the high vowels. The Friedman statistical test showed a significant effect for previous vowel followed by /n/ (X^2 (2, N=10) = 9.455, p = 0.009), and the Wilcoxon tests yielded the following results: (a) for the pair 'mid vowel' vs. 'high vowel' the difference was not significant (Z = -.862, p= .389); (b) for the pair for the pair 'high vowel' vs. 'low vowel' the difference was significant (Z = - 2.198, p= .028); and (c) for the pair 'low vowel' vs. 'mid vowel' the difference was also significant (Z = -2.552, p= .011). The results indicate, thus, that there is an effect of a low previous vowel to favor the accurate identification of the English syllable-final nasal /n/. These results corroborate those found by Kluge et al. (2007) as hypothesized; however this effect of the low previous vowel was not found in the identification of the English coda nasal /m/.

4. Conclusion

This study hypothesized that the Audio/Video condition would favor the accurate identification of the English syllable-final nasal. Results supported this hypothesis, but only in comparison to the Audio only condition, as there was a significant difference between the results of the Audio/Visual and the Audio only condition; for both English syllable-final nasals /m/ and /n/. Results indicate that the Brazilian learners seemed to benefit from the AV presentation in the accurate identification of the target English nasal consonant. Results also showed a slight tendency for the Audio only condition to disfavor the accurate identification of both bilabial and alveolar nasal consonants compared to the AV condition. In summary, the Brazilian EFL learners seemed to benefit from AV presentation as discussed by Grant and Seitz (1998), in the accurate identification of the English nasal consonant /m/ and /n/ in syllable-final position. These results confirm the prediction that Brazilian learners of English would benefit from the AV presentation as the labial/alveolar contrast investigated is a visually distinctive contrast.

As regards the influence of phonological context, results showed that previous vowel influenced the accurate perception of English coda nasals for both nasal consonants. Results also revealed that low previous vowel favor the identification of the alveolar coda nasal and high previous vowel disfavors the accurate identification of the English coda nasal /m/.

Acknowledgements

The first two authors would like to thank the Committee for Postgraduate Courses in Higher Education (CAPES) and the fourth author would like to thank the National Council for Scientific and Technological Development (CNPq), both from the Brazilian Ministry of Education, for funding the present research.

References

Câmara Jr, J. M. (1971). *Problemas da lingüística descritiva.* Petrópolis: Editora Vozes.

Cristófaro Silva, T. (1999). *Fonética e fonologia do português: Roteiro de estudos e guia de exercícios*. São Paulo: Contexto.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, *95*(6), 3623-3641.

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. 233-272). Timonium, MD: York Press.

Fujimura, O., & Erickson, D. (1997). Acoustic phonetics. In W. J. Hardcastle, & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 65-115). Cambridge, MA: Blackwell Publishers.

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

Hazan, V., Sennema, A., Ortega-Llebariad, M., 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*, 360-378.

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

Kluge, D. C., Rauber, A. S., Reis, M. S., & Bion, R. A. H. (2007). The relationship between perception and production of English nasal codas by Brazilian learners of English. In *Proceedings of Interspeech 2007* (pp. 2297-2300). Antwerp, Belgium.

Kurowski, K. M., & Blumstein, S. E. (1987). Acoustic properties for place of articulation in nasal consonants. *Journal of the Acoustical Society of America*, *81*(6), 1917-1927.

Kurowski, K. M., & Blumstein, S. E. (1995). Acoustic properties for the perception of nasal consonants. In M. K. Huffman, & R. A. Krakow (Eds.), *Nasals, nasalization and the velum* (pp. 197-222). Academic Press, Inc.

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

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

Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. B. Pisoni, & E. R. Remez (Eds.), *The handbook of speech perception* (pp. 51-78). Blackwell Publishing.

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