The impact of visual cues and lexical knowledge on the perception of a non-native consonant contrast for Colombian adults

Michele Thompson & Valerie Hazan
Speech, Hearing and Phonetic Sciences, UCL, London, UK

m.thompson@ucl.ac.uk, v.hazan@ucl.ac.uk

ABSTRACT

The study investigates the impact of visual cues and lexical knowledge on the identification of a non-native phonemic contrast. Twenty native Colombians were tested on an identification task involving 16 minimal pairs of English words, produced by four English speakers, contrasting in the presence of /b/ or /v/ in initial or medial position. The test was run in three conditions: audiovisual (AV), audio only (A) or visual only (V). Prior to the identification task, their knowledge of the lexical items was evaluated; they were also recorded while reading the words. Mean identification scores were higher for the AV than the A condition, but V and AV scores not differ. Relative to previous /b/-/v/ studies with Peninsular Spanish speakers, Colombians relied more heavily on visual cues in their identification of /b/-/v/. Although there was a trend for identification scores to be higher for known lexical items, this effect was not statistically significant. Finally, production accuracy for the /b/-/v/ contrast was not correlated with perception accuracy, but production tended to be more accurate in speakers with better lexical knowledge. The visual weighting results suggest that the degree of visual bias in speech perception may be 'culture-specific' rather than merely 'language-specific'.

Keywords: visual cues, lexical effects, L2 perception

1. INTRODUCTION

Face-to-face communication is rich in multisensory information, with both auditory and visual cues available to the listener. Weighting of visual cues typically increases where the auditory channel is degraded (e.g., Sumby & Pollack, 1954). As listening in an L2 can be considered as a form of auditory degradation, it is of interest to consider how second language learners use visual cues when perceiving speech?

Investigation of Japanese speakers' ability to identify contrasts through manipulation of the McGurk effect was conducted by Sekiyama and colleagues (Sekiyama et al, 1993, 2003, 2008) and it was established that Japanese speakers showed less of a visual bias than native English speakers. This may in part due to the fact that visual cues are less informative in some languages, as languages that comprise a greater number of visemes may be more salient for speech-reading (Sekiyama & Burnham, 2008; Hazan et al, 2005). Furthermore, there may also be a cultural perspective to consider as individuals may learn to rely more on auditory cues in cultures where it is considered disrespectful to look at the speaker (Sekiyama & Tohkura, 1993). Hazan et al (2006) explored the effect of visual cues on the perception of non-native phonemic contrasts for Spanish and Japanese listeners in audio visual (AV), auditory-only (A) and visual-only (V) conditions. The contrast chosen was the labial-labiodental contrast, which does not have phonemic status in either Spanish or Japanese. For both the Spanish and Japanese groups, superior performance was obtained in the AV condition followed by the A condition and then the V condition. Spanish listeners' performance was superior in all conditions, with significantly better identification in the V condition than the Japanese listeners. Hazan et al. concluded that familiarity with a visual gesture in the native language may impact on the acquisition of the L2 phoneme; indeed, Spanish speakers are exposed to the labiodental gesture in their native language even though it is not contrastive with the labial gesture, whereas Japanese speakers do not.

Other recent research has focused on the question of whether lexical effects influence the ability to discriminate non-native phonetic contrasts. Mora (2005), in a study with Spanish and Catalan learners of English; significantly higher identification scores were obtained for words than non-words for advanced learners of English, suggesting that lexical knowledge is a significant factor in perceiving non-native phonemic contrasts. Hayes-Harb (2007) compared the effect of phonetic training using minimal pairs (lexical

information) or statistical learning (phonetic information) in a group of monolingual English speakers and reported that both lexical and statistical learning can contribute to the acquisition of a non-native contrast.

The aim of this study was to investigate how Colombian speakers make use of visual and lexical information in their identification of a non-native consonant contrast. The phonemes chosen were the voiced bilabial plosive /b/ and the voiced labiodental fricative /v/. In Colombian Spanish, /b/ is produced as a voiced bilabial plosive unless it is intervocalic or in the post-nuclear position (Moreno & Marino, 1998) when it is produced as the voiced bilabial fricative /β/. /v/ is often substituted for /b/ in the Spanish vernacular but is never found in phonemic contrast with it. It also exists as an allophone of /f/ before voiced consonants.

It is of particular interest to test Colombians as they are reputed to make greater use of eye contact than speakers from Spain. If a similar weighting of A and V cues is found as in the Spanish cohort in Hazan et al. (2005), this would suggest a language effect on the weighting of visual cues (given their shared L1), whereas if they show a greater weighting of visual cues, this would indicate a cultural influence on the weighting of visual cues in speech perception.

The research questions driving this study were as follows:

1. Do Colombian speakers make significant use of visual cues in disambiguating a non-native contrast?
2. Is the ability to discriminate a word pair related to whether the items are known or unknown?
3. Is the ability to produce the novel contrast related to the Colombian speakers' ability to perceive it?

2. METHODOLOGY

The aim was to examine how native Colombians identify the English consonants /v/ and /b/ in three conditions: audio-visual (AV), audio only (A) and visual only (V) when they are in word-initial and word-medial positions. The experiment comprised four sections: (1) completion of an English language questionnaire, (2) translation of the words used in the experiment to establish lexical knowledge, (3) a speech perception identification test, (4) speech recordings of the words used in the experiment. All participants received a consent form and instructions translated into Spanish.

2.1. Participants

Twenty Colombian speakers (15 women and 5 men) were tested in Bogota. Their level of English proficiency was estimated from their responses to an English language questionnaire and an informal evaluation of their speech by the first author, a trained EFL teacher. 55% of the participants were classified as being of low proficiency and 45% as being of intermediate/advanced proficiency. The participants were aged between 19 and 60 years (M=35.2 years; SD 14.5). All had received some basic level of English teaching at secondary school. Five native speakers of British English (4 women, 1 man) served as the control group. They were aged between 22 – 56 years (M=37 years; SD 13.2).

2.2. Materials

Sixteen minimal pairs contrasting /v/ and /b/ were chosen. The word-initial pairs were: ballet-valet, bat-vat, bend-vend, best-vest, bigger-vigour, boat-vote, bowel-vowel, burble-verbal. The word-medial pairs were: cupboard-covered, dribble-drivel, fibre-fiver, hobble-hovel, marble-marvel, rebel-revel, sabre-savour.

2.2.1. Computer-controlled speech perception task

Four speakers (2 women, 2 men) with a Southern British English accent each recorded 32 words on video for a previous study (Hazan et al, 2005), giving a total of 356 items across the three conditions (unfortunately some of the individual tokens were missing resulting in the loss of 28 tokens). Stimuli for the A condition was generated by stripping out the audio track from the video recording (leaving a blank screen for this part of the task), and the V condition was generated though the removal of the audio channel on the video. The listening experiment was designed using DMDX software (Forster, 2002). Two different orders of

presentation of the three modalities were counterbalanced across participants: AV, A, V and A, AV, V. Within each test condition, tokens were randomised across speakers.

2.2.2. Identification of Colombian participants' productions by native English speakers

The eight minimal pairs that were the most reliably produced by Colombian participants were chosen for the identification test with native English speakers. Productions of these 16 words by each participant were digitised and intensity-normalised. They were then presented in an identification test designed using Praat software (Boersma & Weenink, 1997). Ten native English speakers who had no knowledge of Spanish, performed the task. After hearing each word produced by a Colombian speaker, they had to decide whether the word contained a /b/ or a /v/ by clicking on either the B or V label on the screen.

2.3. Test procedure

After completing the English language questionnaire and translating the stimuli, The Colombian participants performed the computer-controlled identification task. This was presented on a laptop with the stimuli presented at a comfortable listening level through headphones. After hearing or seeing each word, they had to click on the label B or V to indicate which consonant they heard. This part of the test was approximately 35 minutes in duration. During the final part of the experiment, a Sanyo Digital Talkbook (ICR-B80NX) was used to record the participants while they read out the test words from a printed sheet.

3. RESULTS

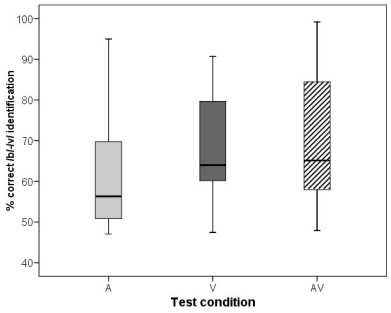
3.1. Identification Task – Control Participants

Mean scores for the five native controls were: A condition: 97.5 % (SD 1.3); V condition: 94.2% (SD 2.0) and AV condition: 99.2% (SD 1.19). T-tests showed a significant difference in performance between AV and V conditions ($t(8) = 4.69$, $p = 0.002$) and A and V conditions ($t(8) = 2.98$, $p = 0.017$), but that the A and AV conditions did not differ ($t(8) = -2.15$, $p = 0.064$). These data suggests that the /b/-/v/ contrast is very salient visually for native speakers.

3.2. Identification Task – Colombian Group Results

The mean scores across the three conditions (see Figure 1) were: A 62.7% (SD 14.9); V 68.1% (SD 13.1); AV 70.0% (SD 16.8). There was a great deal of individual variability in this task: mean identification scores across all three conditions ranged from chance (48.9%) to near ceiling level (94.4%). A repeated-measures ANOVA with within-subjects factors of condition (A, AV and V) and position (word initial, word medial) and between-subjects factor of proficiency indicated a significant main effect of condition ($F(2,36) = 7.013$, $p = 0.003$, $\eta_p^2 = .280$). Paired comparisons with Bonferroni adjustments showed that the condition effect was due to better performance in the AV and V conditions (which did not differ from each other) compared to the A condition. The main effect for position was also significant ($F(1,18) = 26.77$, $p < 0.001$, $\eta_p^2 = .598$) with higher identification rates for the word-initial tokens (M=71.2%) than the word-medial tokens (M=63.5%). There was no significant interaction between condition and position. The effect of language proficiency was not significant: proficiency in English for this group therefore had little impact on the perception of the /b/-/v/ contrast.

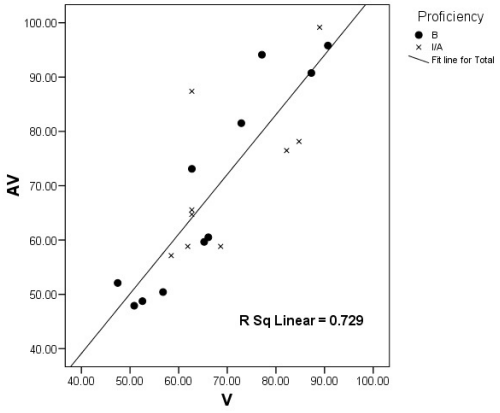
Figure 1: Box plot to show the proportion of correct responses (%) per condition



3.3. Correlations across conditions

Pearson's product-moment correlations were run to see whether identification using auditory cues was correlated with identification when the contrast was cued visually. An R^2 of 0.55 was obtained for the correlation between A and V scores, of 0.69 for the correlation between A and AV scores and of 0.73 for the correlation between V and AV scores. Performance by Colombian participants when only visual cues were presented was therefore strongly correlated with their AV scores. As shown in Figure 2, few participants obtained a higher score in the AV condition than they did when only visual cues were presented. These results demonstrate the participants' strong visual bias.

Figure 2 : Correlation between identification scores in the V and AV conditions for Colombian participants classified in terms of their language proficiency (beginner, Intermediate/Advanced)



3.4. Lexical effects

The word translation task was used to provide a measure of word knowledge per participant. Words that the participants did not know were treated as non-words without lexical representations, requiring bottom-up processing. Words known by the participants were assumed to have some form of lexical representation. By comparing identification scores for known and unknown tokens, it is possible to get a sense of the effect of word knowledge on identification for each participant. There was no significant difference in identification scores for 'known' words (M = 67.5%, SD 21.8) and 'unknown' words (M= 64.6%, SD = 14.56) although there was a trend for lower scores for unknown words. Interestingly, some of the unknown words yielded high scores, e.g. the mean identification score for *hobble* (meaning unknown to 100% of participants) was 82%.

3.5. Production scores

Production scores were obtained by calculating the percentage of /b/-/v/ productions for each participant that were correctly identified by the ten English listeners. Intermediate-Advanced speakers of English (M = 69.6%,) were significantly more accurate in their production than the Beginner group (M = 59.1%) (F (1,18) =8.27, p = .010), although there was a great degree of individual variability in both proficiency groups (See Fig. 3). The effect of position was not statistically significant (F (1,19) = 3.72, p = .07, n.s.).

Figure 3: Box plot of mean production scores per word position and level of proficiency

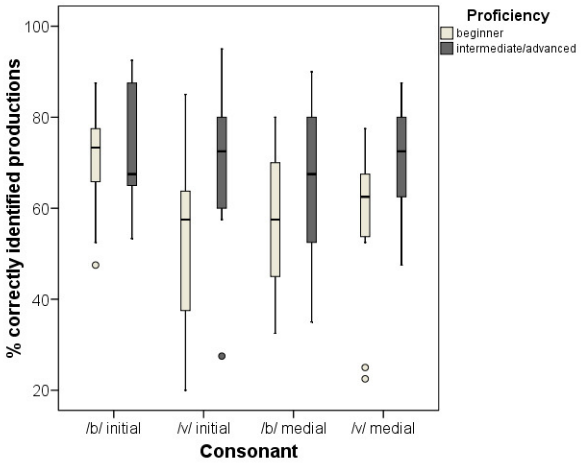
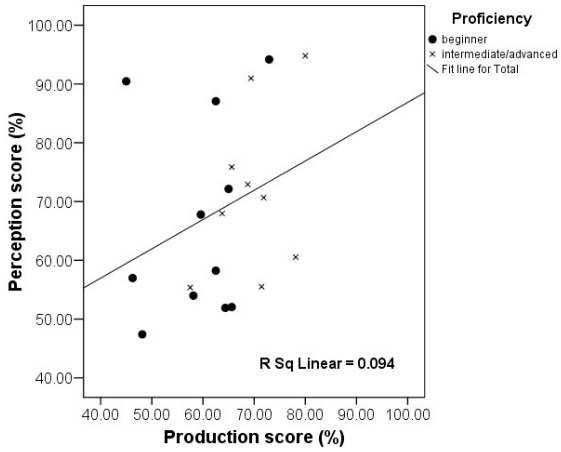


Figure 4 shows that production scores (i.e., how accurately the Colombians' productions of /b/-/v/ were perceived by native listeners) and their perception scores for the same 16 words were not significantly correlated.

Figure 4: Correlation between perception and production scores



3.6. Production and word knowledge results

Correlations between participants' word knowledge (mean percentage of correctly translated words) and their production accuracy was examined for word-initial and word-medial stimuli. There was a significant correlation between these two measures for /b/-/v/ in word-initial position ($r=0.479$; $p<0.05$): participants with better lexical knowledge of the words in the word-initial category tended to produce word-initial /b/-/v/ with greater accuracy.

4. DISCUSSION

This study found that visual cues were of great importance to the Colombians in their identification of the non-native /b/-/v/ contrast. Whereas previous studies with Peninsular Spanish speakers using the same contrast had found that there was no statistical difference between A and AV conditions, showing a greater weighting given to auditory cues (Hazan et al, 2006), our results with Colombian speakers showed that they did not show any difference between V and AV conditions, and that scores for the V and AV conditions were superior to scores in the A condition. As both Peninsular and Colombian speakers share the same L1 phoneme inventory, this result gives some support to the view that the weighting of visual cues in speech perception can be at least partly determined by cultural factors such as the tendency to use eye contact in speech communication. Some support for this view of increased eye contact in Colombian speakers comes from Norris's single case study (Norris, 2007) which states that "the differences lie in the modes of proxemics and gaze, for example, a member of the Colombian discourse system takes up a closer distance and engages in more direct eye contact than a member of the Hispanic discourse system". Cultural factors were suggested in previous studies (Sekiyama and Tokhura, 1993) as an explanation for the reduced weighting of visual cues in speakers of countries like Japan, where there is a culture of 'gaze avoidance'.

With regard to lexical effects, results showed that word knowledge was not a significant predictor in the identification of the non-native contrast but that 'known' lexical items tended to be more accurately produced than unknown words, at least in word-initial position. Therefore the ability to perceive a novel contrast is not always governed by lexical representations. In this study even the participants with low proficiency in English discriminated novel contrasts without referring to the lexicon, albeit inconsistently. Rather, perception seems to rely on a combination of factors namely the relationship between the L1 and L2 sound systems and an interaction between top-down and bottom-up processing mechanisms.

Finally, our study replicated previous findings (e.g., Bradlow et al., 1997) of weak correlations between the ability to perceive and produce a novel phonemic contrast. In both proficiency groups, there was evidence of speakers who could perceive the contrast accurately but could not produce it consistently, and of speakers with accurate productions but perception around chance level.

5. REFERENCES

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

Boersma P., Weenink D. 1997. Praat. Retrieved 12th January 2010 from University of Amsterdam. Website: http://www.fon.hum.uva.nl/praat/

Forster K, 2002. DMDX Display Software. Retrieved 18th June 2009 from University of Arizona, Department of Psychology. Website: http://www.u.arizona.edu/~jforster/dmdx.htm

Hayes-Harb R., 2007. Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research* 23 (1), 65 – 94.

Hazan V., Sennema A., Faulkner A., Ortega-Llebaria 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.

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.

Mora, J. C. 2005. Lexical knowledge effects on the discrimination of non-native phonemic contrasts in words and nonwords by Catalan/Spanish bilingual learners of English. *Proceedings of the ISCA Workshop on Plasticity in Speech Perception*, 43–46. London: University College London

Moreno A., Marino J.B. 1998. *Spanish Dialects: Phonetic Transcription. Proceedings ICSLP 1998*, paper 0598

Norris S. 2007. The micropolitics of personal national and ethnicity identity. *Discourse & Society,* 18 (5). 653 – 674.

Pallier C., Colome A., Sebastian-Galles N. 2001. The influence of native-language phonology on lexical access: exemplar-based vs. abstract lexical entries. *Psychological Science,* 12, 445–49

Sekiyama K. & Burnham D. 2008. Impact of language on development of auditory-visual speech perception. *Developmental Science,* 11 (2), 306 – 320.

Sekiyama K,, Burnham D., Tam, H., Erdener D. 2003. Auditory-visual speech perception development. Japanese and English speakers. *Proceedings AVSP 2003*, 43 – 47.

Sekiyama K, Tokhura, 1993. Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics,* 21, 427 – 444.