Proceedings of Meetings on Acoustics

Phonetic and orthographic cues are weighted in speech sound perception by second language speakers.

--Manuscript Draft--

Manuscript Number:	POMA-D-16-00013R1
Full Title:	Phonetic and orthographic cues are weighted in speech sound perception by second language speakers.
Article Type:	ASA Meeting Paper
Corresponding Author:	Anastasia Giannakopoulou, PhD University of Bedfordshire Luton, Bedfordshire UNITED KINGDOM
Order of Authors:	Anastasia Giannakopoulou
	Maria Uther
	Sari Ylinen
Abstract:	Second language (L2) learning can involve processing of speech-sound contrasts that have multiple phonetic cues (e.g. Iverson et al., 2003). This can be particularly difficult for foreign-language learners especially if the cues are weighted differently in the foreign and native languages (e.g., Giannakopoulou et al., 2011, 2013). The orthographic representation of words is suggested to also interfere with speech sound perception in way of presenting additional cues for the second language learner. Greek child and adult speakers of English were studied to determine on what basis they are making perceptual identification between English vowels with the use of pictures as visual stimuli. Performance was impaired for Greek speakers across all tasks but worst for Greek speakers for the picture stimuli task. Findings suggest a 'link' between orthography and perceptual identification serving as an additional cue for L2 speakers.
Section/Category:	Speech Communication

166th Meeting of the Acoustical Society of America

San Francisco, California
02-06 December 2013

Speech Perception: Paper 2aSC4

Phonetic and orthographic cues are weighted in speech sound perception by second language speakers: Evidence from Greek speakers of English.

Anastasia Giannakopoulou
Department of Psychology,University of Bedfordshire, Luton, UK;anastasia.giannakopoulou@beds.ac.uk

Maria Uther
Department of Psychology, Winchester University, Winchester, UK; maria.uther@winchester.ac.uk

Sari Ylinen
Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland;sari.ylinen@helsinki.fi

Second language (L2) learning can involve processing of speech-sound contrasts that have multiple phonetic cues (e.g. Iverson et al., 2003). This can be particularly difficult for foreign-language learners especially if the cues are weighted differently in the foreign and native languages (e.g., Giannakopoulou et al., 2011, 2013). The orthographic representation of words is suggested to also interfere with speech sound perception in way of presenting additional cues for the second language learner. Greek child and adult speakers of English were studied to determine on what basis they are making perceptual identification between English vowels with the use of pictures as visual stimuli. Performance was impaired for Greek speakers across all tasks but worst for Greek speakers for the picture stimuli task. Findings suggest a 'link' between orthography and perceptual identification serving as an additional cue for L2 speakers.

1. INTRODUCTION

The formation of native-like representations of second-language (L2) phonetic categories is a difficult task for the L2 learner, especially when phonetic cues are used differently between the native language (L1) and the L2 (e.g. Ylinen *et al.*, 2010; Cutler & Broersma, 2005). There are various kinds of cues that need to be weighted and integrated in order for a speech sound to be correctly identified (Holt & Lotto, 2006). Native speakers of a language learn to weight cues in way that those cues carrying more critical information are weighted higher than others (Holt & Lotto, 2006). A second-language (L2) learner's native language (L1), however, may inhibit correct cue weighting in the L2 (Bohn, 1995; Strange, 1998).

L2 speech sound learning is based on the degree of perceived cross-language phonetic similarity: either due to the L2 learner's discrimination of the phonetic differences of two L2 sounds, or due to the existing phonetic differences between L1 and L2 sounds. It is predicted that new L2 sound categories are formed depending on the degree of phonetic dissimilarity between the L2 sound and its closest counterpart in the L1 (Flege, 2003). Iverson *et al.* (2003) argue that the acoustic cues that L2 learners use in practice are often different to those used by L1 speakers, thereby leading to inferior performance on perceptual categorization tasks. The cue weighting mismatch between an L1 and L2 could result in L2 learners identifying and relying on cues that are non-critical for identification of phonetic segments in the L2.

The orthographic representation of words is suggested to also interfere with speech sound perception in way of presenting additional cues for the second language learner. As highlighted in Giannakopoulou et al. (2013), differences in orthographic form may lead L2 learners' attention to non-critical cues, such as vowel duration cues. In the case of the English tense-lax /i:/ - /I/ vowel contrast, the word that contains the long vowel /i:/ is usually spelt with a greater number of letters than its counterpart that contains the short vowel /I/, such as . *sheep* vs. *ship, feet* vs. *fit*. Greek participants (both child and adult groups) seem to rely on duration as a primary cue for perceptual identification tasks (e.g. Giannakopoulou et al., 2011, 2013). Duration is known to have a quantitative feature due to its temporal dimension. This is why vowels are often categorized based on their temporal 'length' and are referred to as long or short vowels (e.g. Hillenbrand, Clark & Nearey, 2001). This may give rise to the idea that participants may assume that more letters means longer vowel duration. Therefore, if Greek speakers tend to rely on duration for vowel categorization (long vs. short), then the number of letters in the word stimulus they are presented with could be an additional 'cue' for this categorization.

In order to control for the possibility that orthographic representation of the word stimuli provides cues for perceptual identification this study explores perceptual identification with the use of pictures as visual stimuli as opposed to word stimuli. This experimental paradigm allowed for perceptual identification of the auditory stimulus excluding any potential orthographic cues since the pictures used did not contain orthographic information (see Figure 1).
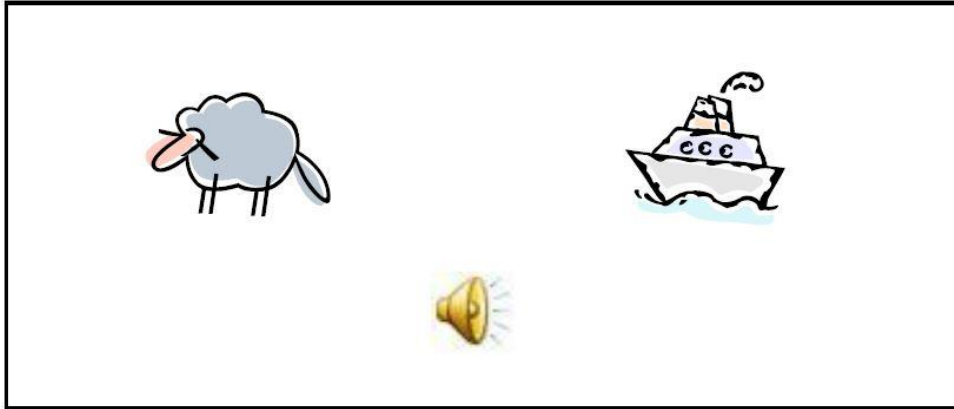
Figure 1: Screen shot of a random picture perceptual identification (PI) task example representing the '*sheep – ship*' word stimulus pair.

Picture naming is of course essentially a translation task, and therefore involves different processes to perceptual identification using orthographic word representation. Picture naming is conceptually mediated (Potter et al., 1984) and involves retrieving the picture's concept. Hence translation inevitably involves concept retrieval as part of the translation process. Kroll & Stewart (1994) used a bilingual translation experiment which included naming and translation experimental conditions. They found that L1 to L2 translation was conceptually mediated whereas L2 to L1 translation was reliably faster and more accurate leading to the proposal that L2 to L1 translation is lexically mediated.

The present study, although it does not test the above assumptions directly, aims at considering perceptual identification processes in the case of L2 picture naming and L2 word translation. Specifically it is hypothesized that: if the auditory stimulus has been conceptually established in the case of Greek L2 English learners, then it may be possible to correctly identify and choose the relevant picture (from the represented pictorial minimal pair). If the previous orthographic representation of the minimal word pair provided some additional cues, then performance in this task could be impaired. Also, in case the picture stimuli were difficult, non-prototypical representatives for the underlying word meanings, or word meanings were un-known, results would also reveal low scores. Considering age group differences, there could be an expectation that children perform worse in general than adults in a reading task due to a slower reading ability (e.g. Balaban, Snidman, & Kagan, 1997), although they might perform better with a picture task given that reading skills are not directly necessary for this task type. On the other hand, given that this is essentially a translation task, it is likely that there could be an effect on children's performance (i.e. giving lower scores) if they have to mentally transform a visual representation to text.

The use of pictures as visual stimuli in this study, thus, required control for the following: 1. Participants' familiarity with the concepts underlying the pictures used; 2. Participants' familiarity with the connection of L2 words and their L1 equivalents (for all the meanings represented by the pictures) which would also control for the degree of any lexical bias/knowledge (e.g. Thompson & Hazan, 2010). This would indicate the degree of participants' access to the semantic inventory as well as access to the conceptual inventory for adult and child groups alike. The main aim of this study is to explore whether Greek learners of L2 English use orthographic cues for auditory perceptual identification, in relation to duration cues (available or non-available) in the auditory stimulus.

2. METHOD

A. PICTURE STIMULI

The same list of minimal word pairs used in Giannakopoulou et al. (2013) was used in the current study and a list of respective picture pairs was created. The pictures were clip art files found on the internet via the Microsoft Clip Art Web application. Thirty picture pairs were used (i.e. 60 picture stimuli) for the present study.

B. PARTICIPANTS

Adult groups: Ten native adult speakers of Standard Modern Greek (6 female, 4 male) aged 19-30 (mean age = 24.3) were tested. They had all lived in Greece and had all studied English as L2 in school (public and private education; L2 English education ranged 8-9 years; mean = 8.5 years). Their level of proficiency in English was advanced in listening, speaking, reading and writing. Ten monolingual English native speakers (5 female, 5 male) aged 19-30 (mean age = 21.4) were also tested and served as controls.

Child groups: Ten child native speakers of Standard Modern Greek (5 female, 5 male), aged 7-8 (mean age = 7.8) were tested. They had all lived in Greece and were students of L2 English. Their respective level of proficiency in English was basic in listening, speaking, reading and writing (L2 English education ranged 1-2 years; mean = 1.6 years). Ten child monolingual native speakers of Standard English were also tested and served as controls (6 female, 4 male; mean age = 7.9). They were recruited from local primary schools in the UK. All participants had normal or corrected to normal vision and none reported any history of a speech or hearing impairment.

C. APPARATUS

Stimuli were presented on a laptop (AMD Sempron) with E-Prime software (Schneider *et al*. 2002a,b). All auditory stimuli were binaurally presented through headphones (SONY MDR-V150) at 44 kHz, 16-bit resolution and at a comfortable listening level (varying between 65-75 dB). Visual stimuli were displayed on a 33 x 20 cm monitor. Reaction times (RT) and responses were automatically recorded for each participant through the E-Run software application. Participants responded by pressing a button on the computer keyboard which triggered the next trial after a 1000 ms delay.

D. PROCEDURE

This study included two main tasks: a paper-based task, and a computer-based perceptual identification task. Greek adult and child participants took part in all tasks, English adult and child participants took part in the computer-based tasks only. Computer-based tasks were counterbalanced. Participants were tested individually. Each session lasted approximately 30-45 minutes.

Picture Naming Task and Translation Task (control, paper-based task): Greek participants took part in two short paper-based tasks: they were provided with a printed list of the picture stimuli arranged randomly, and a numbered list of the respective English words also presented in random order on a separate piece of paper. They were asked: 1. to match the English words to the picture; 2. to translate the list of English words into their Greek counterparts.

Perceptual Identification Task (PI) with picture stimuli: This task used 30 natural vowel duration and 30 modified vowel duration minimal pairs of English words arranged into two

conditions (natural and modified duration stimuli[1] respectively) and the respective picture stimuli. Thus, 60 auditory word stimuli were presented for each condition, 120 in total (30 minimal pairs x2 tokens =60 stimuli x2 conditions =120 stimuli). Participants were instructed to choose which one of the two pictures presented on the screen they heard through pressing a relevant key on the computer keyboard. Due to the nature of the stimuli in the modified duration stimuli condition, participants could only base their choice on spectral rather than durational cues to make the vowel identification.

Perceptual Identification (PI) Task with word stimuli (control task): A similar PI task was also administered, using 30 minimal pairs for each condition presented in their orthographic form. No picture stimuli were used in this task.

Ethical approval (Brunel University Ethics Committee) was obtained prior to conducting the study. Informed consent was given by all adult participants and by the parents or carers of child participants. Participation to this study was on a volunteer basis. All participants received stationary items as incentive for their participation.

3. RESULTS

Figure 2 shows the accuracy scores for Greek adult and Greek child groups for the Picture Match task and the Translation task. Adult participants did slightly better than child participants in both tasks. Results were analysed with a one-way ANOVA with Age group (2 levels) as factor. There was a significant main effect of age for both tasks, Picture Match task ($F_{(1, 18)} = 6.918$, $p < .05$) and Translation task ($F_{(1, 18)} = 10.918$, $p < .05$).
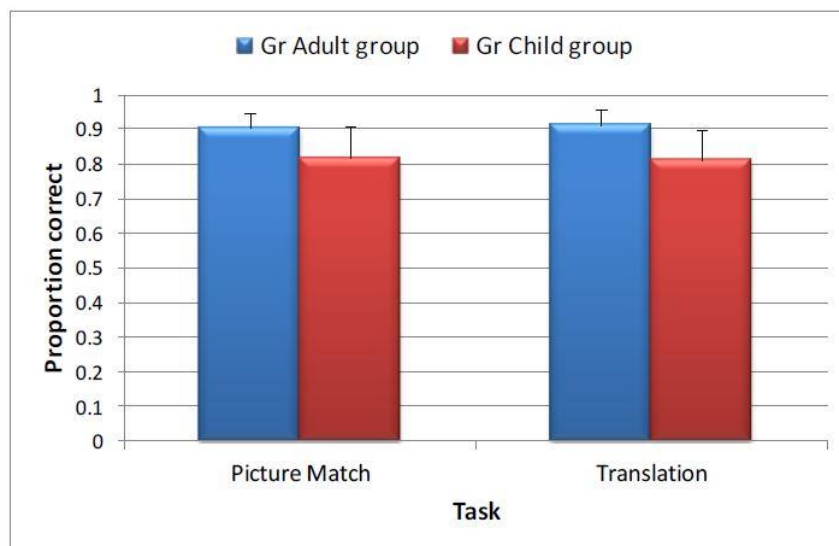

Figure 2: Accuracy scores for Greek adult and Greek child groups for the Picture Match task and Translation task.

---

[1] For a detailed review on natural and modified duration stimuli procedure, refer to Giannakopoulou et al. (2013).
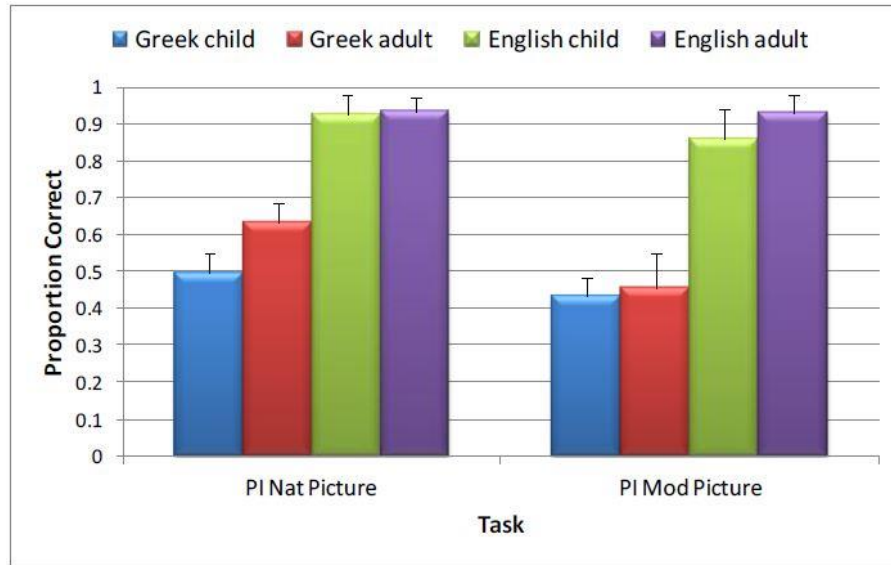
Figure 3: Accuracy scores for adult and child groups for the Perceptual Identification (PI) Picture task for Natural (Nat) and Modified (Mod) duration condition.

Accuracy scores for all computer-based tasks were analysed with a three-way mixed design ANOVA with Language Group and Age as between subject factors and Vowel Length (duration) as a within subject factor. There was a significant main effect of Language (F(1, 36) = 619.895, p< .001), see Figure 3. There was a main effect of Age (F(1, 36) = 17.430, p< .001), see Figure 3. There was a main effect of Vowel-length (F(1, 36) = 56.542, p< .001), see Figure 3. A Vowel-length x Language interaction (F(1, 36) = 15.050, p< .001, see Figure 3) emerged because Greek participants performed better in the natural duration condition than the modified duration condition, whereas English participants showed less marked difference in performance between duration conditions. No other interaction effects were significant.
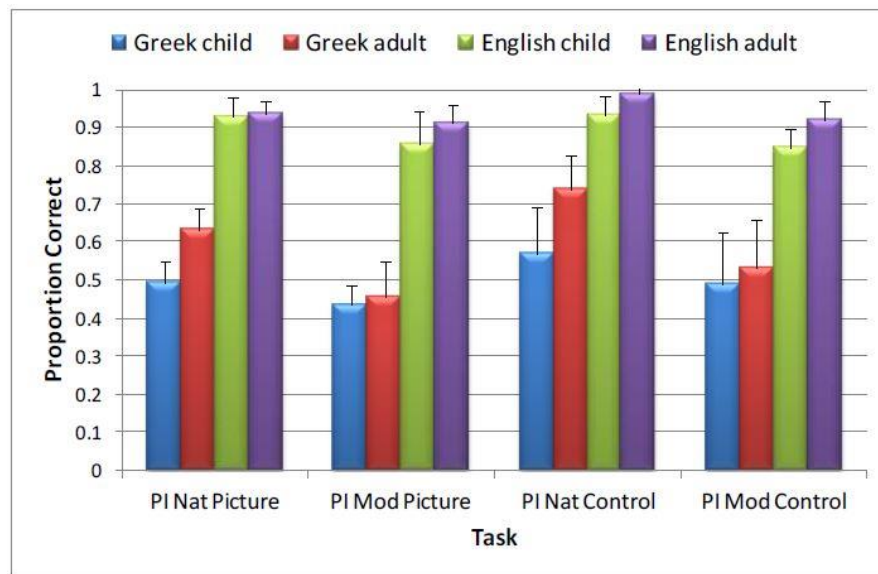

Figure 4: Accuracy scores for adult and child groups for the Perceptual Identification (PI) Picture and Control task for Natural (Nat) and Modified (Mod) duration condition.

English groups performed better than Greek groups across all tasks. The main effect of Language (F(1, 36) = 479.613, p< .001), see Figure 4, was significant, showing that English speakers performed better than Greek speakers overall. The main effect of task (PI Picture task vs. PI Control task) was significant and Greek groups performed worse in the Picture compared to the control PI task (F(1, 36) = 11.110, p< .05, see Figure 4).

4. DISCUSSION

This study explored the degree of reliance on duration cues and orthographic cues. This was tested through the use of perceptual identification tasks, using natural and modified duration auditory stimuli, and the replacement of the orthographic representations of word minimal pairs as visual stimuli with picture pairs. Results revealed high scores on performance with the picture-match task and translation task, with the Greek adult group performing slightly better than the Greek child group. High scores for both tasks indicate familiarity with the concepts and the word meanings used and no effects of lexical bias or knowledge were observed. Kroll & Stewart (1994)'s proposal that L2 to L1 translation is lexically mediated is also supported by the high scores in the translation.

Findings suggest that Greek participants (both adult and child groups) rely on duration as a primary cue rather than spectral cues that are present. This was observed across all task types based on the task conditions used (i.e. natural and modified duration). Performance was lower for Greek groups for all tasks and the modified duration condition in particular. This type of pattern is suggestive of heavy reliance upon duration as a primary cue. This, however, was not the case with English control groups that show ceiling effects across all task types. It could therefore be suggested that Greek participants primarily use the 'wrong' or non-critical cues for perceptually identifying English vowel distinctions and performance drops in the case these cues are not available, irrelevant or misleading. Additionally, results confirm outcomes of Giannakopoulou et al. (2013) that duration is not used by English native speakers as a primary cue for correct perceptual identification, unlike Greek learners of L2 English who show preference to duration cues rather than spectral cues for correct vowel identification.

L2 child participants' low scores in the perceptual identification task compared with L2 adult participants' scores could be attributed to a combination of age and proficiency effects. The fact that L2 child participants had higher accuracy scores for the Picture Match and Translation tasks compared with their lower scores (at chance level) for the perceptual identification task types indicates task bias rather than lexical bias. Age and proficiency level could, however, be contributing factors which should be further tested in future experimental designs in order to identify which of these factors is more likely.

Finally, the fact that the Greek adult group performs better in the PI control task which involved words as visual stimuli compared to the PI task which used picture stimuli (although auditory stimuli were identical for both tasks), suggests that the orthographic representation of words could be providing *additional* cues that were not available in the case of picture stimuli. The low scores (at chance level) by Greek child participants in the perceptual identification tasks and particularly for the picture task, suggest an interference between reliance to orthographic cues and also possible effects of age versus proficiency levels in comparison to Greek adult participants, specifically for the picture and word tasks with natural vowel duration stimuli.

Findings suggest a 'link' between orthography and perceptual identification in L2. Thus, on the one hand, word orthography may facilitate perceptual identification if it works as an additional cue. On the other hand, however, cues provided by orthography may not be the same as those used by native speakers, especially when the orthography is not transparent. In the

present study, for example, the English orthography may have supported L2 speakers' cue weighting that is based on duration, although native English speakers weighted spectral cues. Further research would be necessary in order to extend current findings, as well as further explore the suggestive account of perceptual interference of orthographic cues in the case of perceptual identification and discrimination of speech sounds by L2 speakers. The current findings on the relationship between orthographic and auditory weighting of cues increase the understanding of the strategies used to acquire new languages.

ACKNOWLEDGMENTS

REFERENCES

Balaban, M. T., Snidman, N. & Kagan, J. (1997). Attention, emotion, and reactivity in infancy and early childhood. In P. J. Lang & R. F. Simons (Eds.), *Attention and orienting: Sensory and motivational processes,* 369-391. Hillsdale, NJ: Erlbaum.

Bohn, O. S. (1995). Cross language speech perception in adults: first language transfer doesn't tell it all. In: W. Strange, (Ed.). *Speech perception and linguistic experience: issues in cross language research*. Timonium, MD: York Press. 279–304.

Cutler, A. & Broersma, M. (2005). Phonetic precision in listening. In: W. J. Hardcastle, & J. Mackenzie Beck (Eds.). A figure of speech: A Festschrift for John Laver. London: Lawrence Earlbaum. 63-91.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In: A. Meyer & N. Schiller (Eds.). *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, Berlin: Mouton de Gruyter, Berlin. 319-355.

Giannakopoulou, A., Uther, M., Ylinen, S. (2013). *Enhanced plasticity in spoken language acquisition for child learners: evidence from phonetic training studies in child and adult learners of English*. Research and practice in working with children who are bilingual or have English as an additional language and who have language and communication needs. Special Issue of: Child Language Teaching and Therapy. SAGE Publications.

Giannakopoulou, A., Uther, M., Ylinen, S. (2011). *Phonetic cue-weighting in the acquisition of a second language (L2): evidence from Greek speakers of English. In:* Achievements and perspectives in SLA of speech: New Sounds 2010. Volume 1. (Eds.) Magdalena Wrembel, Malgorzata Kul and Katarzyna Dziubalska-Kolaczyk. Frankfurt am Main: Peter Lang Verlag. Pp. 91-102.

Hillenbrand, J., Clark, M., & Nearey, T. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America*, 109(2), 748–763.

Holt, L. L. & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, 119, 3059-3071.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A. & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47–B57.

Kroll, J. F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33,* 149-174.

Potter, M. C., So, K. F., Von Eckhardt, B. & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior,* 23, 23-38.

Schneider,W., Eschman, A. & Zuccolotto, A. (2002a). E-Prime User's Guide. Pittsburgh: Pychology Software Tools Inc.

Schneider,W., Eschman, A. & Zuccolotto, A. (2002b). E-Prime Reference Guide. Pittsburgh: Pychoogy Software Tools Inc.

Strange, W. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, 26, 311–344.

Thompson, M. & Hazan, V. (2010). The impact of visual cues and lexical knowledge on the perception of a non-native consonant contrast for Colombian adults. *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010*. 493-498.

Ylinen, S., Uther, M, Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R. & Näätänen, R. (2010). Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English. *Journal of Cognitive Neuroscience*, 22 (6), 1319-1332.