Emotional Prosody Mediated Visual Search in a First and Second Language: Evidence from Eye-Movements

Patra Vlachopanou

University of Essex, UK

0586

The Asian Conference on Society, Education and Technology 2013

Official Conference Proceedings 2013

iafor
The International Academic Forum
www.iafor.org

Introduction

In recent years, a growing number of researchers investigated how eye tracking uses eye movements, in order to examine the interaction of perceptual systems which are involved with language and vision. Of particular interest was how language processing can influence visual perception (Spivey et al., 2001). Furthermore, other eye tracking studies had shown that linguistic prosody is used by listeners in order for them to understand and interpret what is said and to produce anticipatory predictions during real-time comprehension of spoken language (Ito & Speer, 2008). The purpose of the present study was to investigate if emotional prosodic cues—the rhythm and intonation of speech and the emotional state of the speaker—could similarly influence visual perception. Interest in whether participants' eye movements could be influenced by emotional prosody in real-time language comprehension has recently come to the forefront in the study of cognitive neuroscience in language by Paulmann et al. (in press). There is evidence that in every culture and language, utterances which are emotionally nuanced, are very important in helping people to understand the emotional meaning that exists behind the phrase. For example, imagine that someone asks how you are and you answer that you are fine. When the word "fine" is spoken in a sad tone of voice, the listener understands that the speaker does not really mean what they say. In fact the listener believes that the phrase conveys another emotional meaning (Sperber & Wilson 1986, cited in Kitayama & Ishii, 2002). The current study went further, investigating if emotional prosody could influence participants' eye movements, while they are listening to emotionally intoned instructions spoken in their first (L1) or second (L2) language during a visual search task. Bilingualism has always been a challenging area for the researchers who wish to explore the depth of bilinguals' minds (Pavlenko, 2006).

The Current Study

For the present study, a modified Stroop task was used. Participants' eye-movements were monitored as they looked at four faces with different emotional expressions on a computer display, while they were listening to emotional instructions uttered in two languages, German (L1) and English (L2). These utterances were either congruent or incongruent with the emotional prosody. To investigate the influence of emotional prosody as a clue for emotional face evaluation, the current study used eye-tracking equipment. A 'pre-time window' that measured the fixations and the accuracy of eye movements before the onset of the emotional adjective, was also used. The hypothesis of the present study was that reaction times would be longer for faces that mismatched the prosody than for faces that matched the prosody, when participants were listening to instructions in their first (L1) and second (L2) languages. It was also expected that participants would fixate longer on faces that matched the prosody than on faces that mismatched the prosody. Therefore, the study was an investigation into how the tone of voice might influence perception of facial expressions. How long would it take for participants to recognize the facial expressions? Would they use the information they heard or not? Would the fact that the sample was made up of late bilinguals affect their performance when they heard the intoned instructions in their first and second languages?

Method

Participants

Seventeen native speakers of German with English as a second language participated in the current study. They are comprised of 14 females and 3 males, with a mean age of 28.7 years and a mean length of education of 17.4 years. They were recruited through an advertisement that was posted at the University of Essex. All the participants had normal hearing and normal or corrected-to-normal vision. They all completed a consent form before the start of the study, which was ethically approved by the University of Essex. The participants were five undergraduate, ten postgraduate, one University staff member, and one senior teaching fellow. Fifteen of the participants came from Germany, one was from Austria, and one was from Somalia. Participants' mean number of years of education in the U.K. was 2.7 and their mean age of English language acquisition was 11.2 years. Both language preference and language proficiency ratings were obtained for each participant. Furthermore, all participants were asked about their language skills and preference at the beginning of the experiment. Six participants indicated no language preference, five participants indicated that English was their preferred language of communication while three indicated that German was their preferred language to speak in but English was their preferred language to listen to. One participant indicated a preference to speak in the second language and listen in the native language, whereas another participant indicated no language preference in speaking but preferred to listen in the second language. The final participant indicated a preference for the native language for speaking, but had no preference between the languages for listening. The participants were also asked to rate their language proficiency, speaking proficiency, and language independence on a scale of one to ten. The results are presented as follows: participants' mean rating for language proficiency was 9.4, their mean rating of speaking was 8.8, their mean rating for language independence was 9.1, and their mean rating for ease of language switching was 7.7.

Apparatus

Eye movements were recorded by the EyeLink 1000 Core System, consisting of a standard camera which provides an image of the eye that is being tracked. The host PC was a Macintosh computer on which the participants were able to watch the experiment, while their behavioral responses were recorded using a mouse. EyeLink Software was used to fixate, calibrate, and validate the participants' gaze, and finally to run the experiment. Speech files containing the instructions were recorded and played on the host PC. There were two speakers in the experiment: one English and one German native speaker. Both speakers were female.

Materials

Questionnaire. A language questionnaire was first given to participants with questions about their native country as well the number of years of education in their native or other country. The questionnaire also asked about their second language speaking proficiency and grammar skills, as well as language preference and language independence (see appendix).

Auditory Stimuli. The speech stimuli were simple auditory instructions uttered in English or German (‘Click on the face with the xx expression’, ‘Klick auf das gesicht mit dem xx ausdruck’, where ‘xx’ was an emotional adjective). The expressions in English were pronounced by a native English speaker and the expressions in German by a native German speaker. Five different adjectives were used in this experiment, reflecting four different emotional categories (anger, fear, happiness, and sadness), as well as a neutral emotion. The participant heard an auditory instruction to click on a specific face, which corresponded to each of the emotional categories. The instruction was pronounced with different emotional tones of voice (5 target emotions x 5 prosodic emotions x 2 languages=50 auditory stimuli). The utterances used were, ‘klick auf das gesicht mit dem freudigen/aergerlichen/ aengstlichen/traurigen/neutralen ausdruck’ and in English the corresponding request were ‘Click on the face with the happy/angry/frightened/sad/neutral expression’. The utterances were pronounced in a happy/angry/frightened/sad/neutral tone of voice. Pell et al. (2009a, 2009b) had used these specific terms in a previous study, and these results were considered worth mentioning by Paulmann et al. (in press). For that reason, this study used the same emotional terms. At this point it is important to mention that for each emotional category, the speaker produced a sentence in which the semantic content and the prosody matched (e.g. ‘Click on the face with the angry expression’, said in an angry tone of voice), and a sentence in which the semantic content and the prosody mismatched (e.g. ‘Click on the face with the angry expression’, said in a happy tone of voice or another tone except angry). The instructions given to participants were to click on the face which corresponded to the simultaneous instruction that was given from the computer, irrespective of the emotional prosody. Participants were asked be as accurate as possible.

There were pitch variations across utterances. Generally, the angry utterances were spoken with louder and more highly pitched voices, while sad utterances were usually expressed with a lower pitch. Moderate variations in pitch were usually produced for happy utterances.

Visual stimuli. Four black-and-white photographs of 170 x 220 pixels with static facial expressions of an actor’s face were presented to participants. For the presentation of these facial expressions, eight actors were used, four female and four male, of different ethnicities (Caucasian, Black, and Asian). Similar visual stimuli have been used in a previous study with success (Paulmann & Pell, 2010). For that reason, it was thought that it would be wise to use this kind of stimuli again. Each actor presented one exemplar for each of the five emotional categories. In 160 main trials (the practice trials are not included here) participants were presented with 32 angry, 32 happy, 32 sad, 32 frightened and 32 neutral expressions. These emotional expressions were used as they are universally recognized (Ekman, Sorenson, & Friesen, 1969), and there is also evidence that justifies the above statement, from a previous study (Young, Rowland, Calder, Etcoff, Selt, & Perrett, 1997, cited in Paulmann et al., in press).

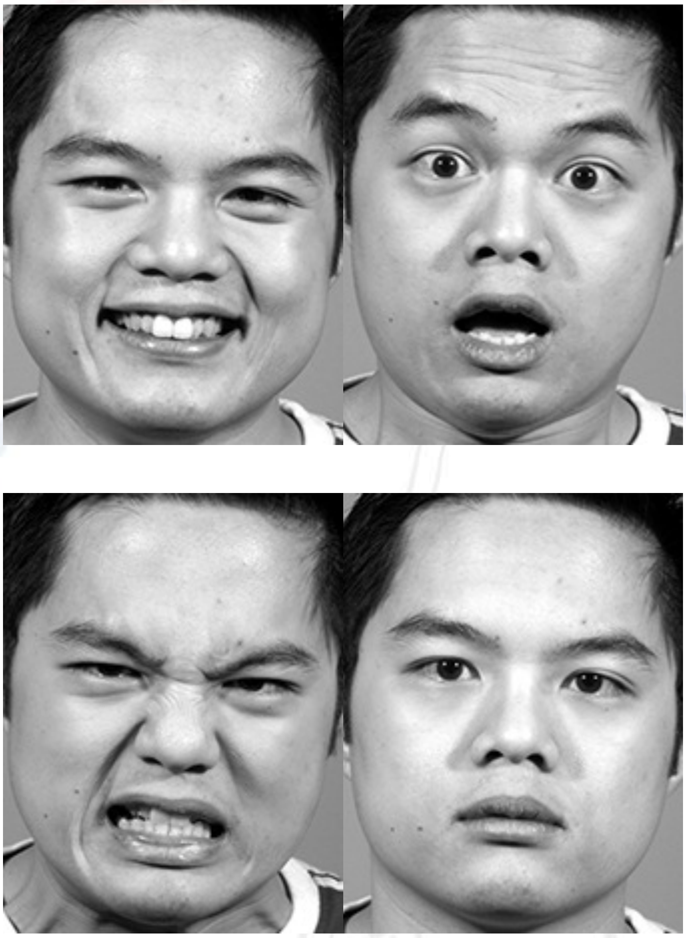

Figure 1. Example of face targets, as presented in the experiment, posed by one of the actors.

Design

Five practice trials were presented at the beginning in order to help familiarize the participants with the experiment, and 160 trials in total represented the actual experiment. Four photographs of the same actor were shown on the display. Each photograph represented a different facial expression. These five expressions were of the target emotions, anger, fear, happiness, sadness, and neutral. The neutral face was always presented on the screen, as well as one face which matched the prosody, and one which matched the semantics. The commands, as mentioned above, were given to the participants in both English and German. Eighty trials were presented in the English language. In forty trials the emotional prosody and the target adjective matched (e.g. 'Click on the face with happy expression', spoken in a happy prosody and the same for the other emotions). In the other forty trials the emotional prosody and the target adjective mismatched (e.g. 'Click on the face with sad expression', spoken in a happy voice). The same procedure was carried out in the trials presented in German. There were eighty trials in this language, as above, with forty where the emotional prosody and the target adjective matched, and forty where the emotional prosody and the target adjective mismatched. A 'match' trial was presented eight times for each emotional prosody (that is, 8 x AN_AN 'Click on the face with the angry expression' with an angry emotional tone), and a mismatch trial twice (that is, 2 x AN_FE 'Click on the face with angry expression' in a fearful emotional tone, 2 x AN_HA 'Click on the face with angry expression' in a happy emotional tone, 2 x AN_NE 'Click on the face with angry expression' in a neutral emotional tone, 2 x AN_SA 'Click on the face with angry expression' in a sad emotional tone = 8 mismatched in total for angry prosody). Two separate 2 x 2 x 5 ANOVAs (language x congruency x emotion) were conducted in a within-subjects experiment. The two dependent variables were the reaction times, meaning how long it took the participants to click on the stimulus face, and the accuracy rates, meaning how accurate participants were in choosing the face that was congruent. Three conditions were taken into account: the match/ mismatch between emotional adjective and face, the English and German languages, and the five levels of emotional prosody (anger, fear, happy, sad, and neutral). The interaction was therefore 3-way, among language, congruency, and emotion.

Procedure

After the preparation of the eye-tracking recording, the participants were seated in a dimly lit room 75 cm away from the monitor. EyeLink 1000 with 1000Hz sampling rate was used to record the eye-movements of participants. Furthermore, participants were instructed to sit straight, to place their head on the chin rest facing the centre of the screen, and to minimize head movements. They were also instructed to listen to the auditory stimuli carefully, and to use the mouse each time the experiment required it. Thus, each time the participant heard the specific request, for example 'Click on the face with the happy expression', they clicked the mouse, which allowed the experiment to continue. Each trial began with the participants fixating on a small black circle in the middle of the screen. After the fixation, the next step was the calibration and the validation of participant's ability to fixate on the small black circle. After this procedure the experiment was started. At the beginning of the experiment, the participants fixated again on the dot in the middle of the display. When participants clicked the mouse the experiment ran and a circular array of faces appeared on the display. Then, the same array was presented again, but this time followed by auditory instructions, requiring participants to click on a specific face as they heard the instructions. The first array of faces was presented in order to help participants familiarize themselves with these faces, as, when the face stimulus is presented in an unfamiliar view, the recognition of face identity is disrupted (Frischen, Eastwood, & Smilek, 2008).

Discussion

The present study investigated if emotional prosody could mediate visual search in both first and second languages. Specifically, we investigated if emotional prosody could be used by listeners to generate anticipatory predictions during on-line speech, and if that would influence the eye movements of participants when listening to emotionally intoned instructions spoken in their first (L1) or second language (L2) during a visual search task. This is the first study which has used eye-tracking methods to examine this kind of hypothesis. What was generally expected was that participants would have longer to reaction times to faces that mismatched the prosody, than to faces that matched the prosody, as emotional prosody has a bigger influence in the first language (Pavlenko, 2008) and processing in second language is less automatic and less efficient (Birdsong, 2006) and is thus more prone to influence from prosody. Our results were derived from two different sources, behavioral data and eye-tracking data.

Behavioral data gave us important information about the current study, verifying the hypothesis. In relation to reaction times, we found that participants took longer to click when the prosody mismatched with the adjective than when the prosody matched with the adjective. Furthermore, participants clicked more quickly when they heard English instructions than when they heard German instructions. Moreover, participants took longer to click when hearing fearful and sad instructions compared to neutral instructions, and were quicker when hearing angry or happy instructions in comparison to neutral instructions. Based on results in accuracy rates, participants made more mistakes during incongruent instructions (e.g. happy prosody/angry adjective) than during congruent instructions (e.g. happy prosody/happy adjective). However, the accuracy rate error was low, which means that participants were very accurate in their trials. Only when the instructions were congruent, the fearful tone of voice led participants to make more mistakes than when they heard the instructions in a neutral tone of voice
.

Eye-tracking data complemented the findings, and gave us information about how emotional prosody could influence participants' gaze, verifying our expectations. These results allow us to understand where the participants' eyes were looking before the onset of the adjective, when they could rely only on emotional prosody. In general, participants made longer fixations to faces that matched the emotion of the prosody than to faces that mismatched the emotion of the prosody. Furthermore, when the emotional tone was fearful or sad, participants made shorter fixations than when the tone of voice was neutral. Moreover, when the instructions were spoken in English, participants made shorter fixations than when these were spoken in German.

Taken together, our data provides evidence to support the predictions made in the current study. According to the eye-tracking results participants fixated longer onto faces that matched the emotional prosody than onto the faces that mismatched the emotional prosody. Furthermore, regardless of behavioral results, participants used the information they had heard in both their first and second languages, as the accuracy rates showed they were correct in both languages when they were responding to the adjective that they were asked to click on. Reaction times were longer for faces that mismatched the prosody than for faces that matched the prosody. Moreover, as late bilinguals, their performance was not the same in their first and second language, as they had shorter reaction times in their L2 than in their L1. The current results nicely complement previous evidence, and extend the research by giving us information about bilinguals' language processing. However, the data did not give us enough information about how exactly emotional prosody guides bilinguals' gaze in all the stages of language processing. This could be a challenge for future investigation.

Many explanations emerge from the data. As far as the behavioral data is concerned, it is plausible that participants would have longer reaction times to incongruent instructions, because until the onset of the adjective, they were looking at the faces. When the adjective was spoken, they clicked on the correct face. Thus, incongruent faces had greater reaction times than the congruent faces. As for accuracy rates, results showed that participants followed the instructions and accurately clicked on the face requested. Whether or not they were influenced by prosody at the beginning of the trial, in the end they clicked on the right face. Concerning eye-tracking data, it appeared that the participants looked longer at matched faces than mismatched ones, because they were influenced by the emotional tone before the onset of the adjective. Thus, they looked longer at the faces that matched the tone. According to the language findings, participants clicked faster in English than in German. Participants knew that they had to complete a task, and they tried to be as accurate as they could, following the instructions successfully.
Our results go hand-in-hand with several findings on congruency and face processing in the literature.

Interpretation of eye-tracking results

A recent study which referred to monolinguals showed that when the stimuli was congruent, participants fixated longer to matched faces, Paulmann *et al*. (in press). This current study went further, verifying that bilinguals fixate longer when the stimuli is congruent than when it is incongruent. Particularly, the pre-emotional time window showed that participants' eye-movements were more influenced by emotional prosody, as there was no semantic information at that time, and fixated longer to congruent stimuli. This clue leads us to agree with the previous study, which stated that the meaning of a vocal emotional expression exists in a listener's memory before the onset of the semantic instruction, and helps us to see that the emotion of the prosody can be implicitly assessed during on-line spoken language processing (Paulmann et al., in press), even if listeners hear the instructions in two different languages. As mentioned in the introduction, participants more easily recognized the voices which were congruent to the face than those which were incongruent (de Gelder & Vroomen, 2000; Massaro & Egan, 1996; Paulmann et al., in press). In the present study the findings complement the previous data. As such, we could state that emotional prosody influences visual attention just as linguistic prosody does (Ito & Speer, 2008), because it allows us to understand an incoming message, and also influences the way that our eyes move. The fact that combined visual and auditory processing can remove ambiguity and help us understand a facial expression (de Gelder et al., 2006; Paulmann, Jessen & Kotz, 2009) might mean that eye gazes could be influenced by the emotion of a visual stimulus (Paulmann et al., in press) and memory could play a crucial role in that. As Bower (1981) and Niedenthal et al. (1994) stated, emotion can be recognized irrespective of language, due to semantic memory structure (cited in Pell & Skorup, 2008). Previous studies showed that the speed with which participants recognize a face depends on specific emotions. Some emotions emerge more quickly than others (Oehman et al., 2001; Batty & Taylor, 2003; Pell 2005a, 2005b; Paulmann & Kotz, 2007, cited in Paulmann et al., in press), so the duration and the speed of the gaze depends on a specific emotional category. Our study discovered the same, complementing the previous findings, showing that bilinguals perform in the same way as monolinguals and their gazes were influenced in the same way. When the emotional tone was angry, participants fixated for a shorter time than when the emotional tone was neutral and/or happy, verifying the finding that negative faces are recognized faster among neutral faces, or compared to happy faces (Calvo et al., 2006; Eastwood et al., 2001; Juth et al., 2005, cited in Calvo & Nummenmaa, 2008). After noting this, and the fact that participants fixated longer to matching faces, it was observed that they clicked on the faces that matched the prosody in the time period between 267-332ms after speech onset. This means that the initial emotional speech processing started at the first 200ms. Likewise, according to Martin, Shap and Boff, (1993, cited in Paulmann et al., in press) the first fixation starts after 150ms. Thus, it is easily understood that fixations to emotional faces will start to increase after 150-200ms. So, it is assumed that we need 267-332ms to extract prosodic information from faces. However, previous studies found that emotional information is extracted a few milliseconds earlier (Paulmann et al., in press). This has led us to believe that our findings were perhaps influenced by the use of two languages. In summary, a rapid emotional prosody is evaluated quickly and guides our eye movements in a few milliseconds. So, it is possible for us to anticipate events. When someone needs at least 200ms to extract emotional meaning, it follows that the eye-movements will make longer fixations to matched faces.

Interpretation of behavioral results

In behavioral results our findings confirm what previous studies have found. Participants had shorter reaction times to congruent instructions than to incongruent ones. As mentioned previously, this finding shows that participants were influenced by prosody before the semantic information. So, at first they fixated to faces that matched the prosody, but not the prosodic/semantic information. After the onset of the adjective, they matched the prosodic/semantic information, so there were longer reaction times to incongruent stimuli than to congruent stimuli. Our findings have been confirmed by previous studies which have found faster reaction times to congruent instructions (Hernandez et al., 2010; Bialystok et al., 2008a, cited in Bialystok et al., 2009). In general, bilinguals have the benefit of performing well in tasks that involve conflict resolution (Bialystok et al., 2006; Costa et al., 2008, cited in Bialystok et al., 2009). However, we must not forget that the initial instructions that were given to participants were to click on the face that matched the instructions. So, our findings were expected. What it is difficult to prove is that emotional prosody was the reason that made them match the face with the instruction, because as we know, semantic information is powerful. About this, research has reported that semantics are difficult to ignore when participants are told to focus on linguistic prosody. However, emotional prosody could be ignored when participants turn their attention to semantic information (Paulmann & Kotz, 2007; Paulmann et al., in press). Moreover, if we take into account the fact that emotional prosody is highly automatic (Schupp et al., 2004a; Hird & Kirsner, 1998, cited in Paulmann & Kotz, 2007) and non-voluntary (Wambacq & Jerger, 2004, cited in Paulmann et al., in press), we can understand that it was logical for participants' eye-gaze to be influenced by the prosody at the beginning of the utterance, but as the utterance was unfolding, it was logical for the semantic information to be dominant. Likewise, according to Paulmann et al., (in press), emotional prosody is presented in the first few milliseconds after speech onset. However, semantic information takes more time to be processed. As far as accuracy was concerned, the error rates were low. This means that participants clicked on the right face in the end. This finding pleasingly complements the previous research of Scherer et al. (2001), who said that when the face and the voice express the same emotion, accuracy rates increase and reaction times of the listener decrease. This is because, combined visual and auditory emotional contexts resolve the ambiguity that facial expressions may occur (de Gelder et al., 2006; Paulmann, Jessen & Kotz, 2009), and manage to guide and facilitate visual search behavior even though the instructions did not require participants to follow the emotional cues (Paulmann et al., in press).

Interpretation of the difference between the first and the second language

In the current study, bilingual participants were used. So far, from the literature, this was the first time that research has used bilinguals to measure whether emotional prosody could influence eye gaze. Our present data has shown a statistically significant effect between the languages. Participants had faster reaction times in their L2 than in their L1. The accuracy rates were however almost the same. Unfortunately, our data cannot tell if it was emotional prosody that influenced participants to have faster reaction times in their second language. It is well known that bilinguals are more emotional in their first language (Pavlenko, 2008), but sometimes they sacrifice their emotions at the altar of successful completion of what they have to integrate. There is evidence, as mentioned in the introduction, that bilinguals are prone to ignore prosody when emotional information, in our case emotionally intoned instructions, captures their attention (Bialystok et al., 2009). Converging evidence from other studies has supported the conclusion that when misleading, irrelevant or false information comes to the fore, bilinguals have the ability to suppress it, and try to be as accurate as possible in what they were asked to do (Zied & colleagues, 2004, cited in Bialystok et al., 2009). Harris et al. (2005, cited in Sutton et al., 2007) stated that the Stroop type interference effect may vary between the two languages due to the age of acquisition. For late bilinguals, someone would expect the interference effect to be greater in L1. In conclusion, bilinguals sacrificed the performance of the first language for the sake of the second language. As the second language is less efficient (Birdsong, 2006), so it needs more effort to be accurate.
The present work has a number of limitations that should be addressed in future research. Although the pre-time window results suggest that emotional prosody can influence eye gaze during early stages of language processing, its influence on the later stages of processing remains to be investigated. Likewise, we cannot prove that bilinguals are influenced only by the emotional prosody, when they hear instructions in both their languages. As stated above, semantic processing is really powerful, and according to the present data this influenced participants' reaction after the onset of the adjective. Furthermore, the language effect does not show that participants' eye gazes were influenced more when they heard the instructions in L1 or in L2. Obviously, further research is necessary before any firm conclusions could be drawn about how emotional cues are processed in early and late language processing, and how eye-gaze could be influenced when emotionally intoned instructions are given in L1 and L2. A future study could measure the exact time of participants' eye movements during the language processing, at different stages. The more measurements taken the more knowledge could be extracted from them. Lastly, it remains to be seen whether or not the current results are produced by the suppression of the first language to enable the processing of the second language.

Conclusion

Taking everything into consideration, the present study, on the one hand has verified the previous findings, and on the other has added its own mark to the history of bilingual research. In short, the results have verified that emotional prosody is rapid and involuntary, that negative faces can be extracted more easily in comparison to neutral, and that it is easier to recognize emotional cues when the stimuli are congruent. Likewise, the results gave us the knowledge that participants, irrespective of language, are influenced by emotional prosody, especially at the early stages of language processing. They also showed that with eye tracking, real time language comprehension can be measured helping science to make a big step in interpreting the emotional stages of human communication.

References

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Brain Research*, *17*(3), 613-620. Elsevier.

Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual Minds. *Psychological Science in the Public Interest*, *10*(3), 89-129.

Birdsong, D. (2006, February). Dominance, proficiency, and second language grammatical processing. *Applied Psycholinguistics*.

Calvo, M. G., & **Nummenmaa**, L. (2008). Detection of emotional faces: salient physical features guide effective visual search. *Journal of experimental psychology General*, *137*(3), 471-494.

De Gelder, B., Meeren, H. K. M., Righart, R., Van Den Stock, J., Van De Riet, W. A. C., & Tamietto, M. (2006). Beyond the face: exploring rapid influences of context on face processing. *Progress in Brain Research*, *155*, 37-48. Elsevier.

Dromey, C., Silveira, J., & Sandor, P. (2005). Recognition of affective prosody by speakers of English as a first or foreign language. *Speech Communication*, *47*(3), 351-359.

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion.*Perception And Psychophysics*, *63*(6), 1004-1013.

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2003). Negative facial expression captures attention and disrupts performance. *Perception and Psychophysics*, *65*(3), 352-358.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86-88.

Frischen, A., Eastwood, J. D., & Smilek, D. (2008). Visual search for faces with emotional expressions. *Psychological Bulletin*, *134*(5), 662-676.

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition & Emotion*, *16*(1), 29-59.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97-115.

Paulmann, S., & Kotz, S. A. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, *1151*, 107-118.

Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *NeuroReport*, *19*(2), 209-213.

Paulmann, S., Jessen, S., & Kotz, S. A. (2009). Investigating the Multimodal Nature of Human Communication. *Journal of Psychophysiology*, *23*(2), 63-76.

Paulmann, S., & Pell, M. D. (2010). Contextual influences of emotional speech prosody on face processing: how much is enough? *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 230-242.

Paulmann, S., Titone, D., & Pell, M., D., (in press).

Pavlenko, A. (2006). Emotions and Multilingualism. Cambridge: Cambridge University Press.

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition*, *11*(02), 147-164.

Pell, M., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication*, *50*(6), 519-530.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, *37*(4), 417-435.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing Emotions in a Foreign Language. *Journal of Nonverbal Behavior*, *33*(2), 107-120.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of CrossCultural Psychology*, *32*(1), 76-92.

Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, *15*(8), 1135-1148.

Spivey, M. J., Tyler, M. J., Eberhard, K. M., & Tanenhaus, M. K. (2001). Linguistically Mediated Visual Search. *Psychological Science*, *12*(4), 282-286.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643-662. American Psychological Association.

Sutton, T. M., Altarriba, J., Gianico, J. L., & Basnight-Brown, D. M. (2007). The automatic access of emotion: Emotional Stroop effects in Spanish-English bilingual speakers. *Cognition & Emotion*, *21*(5), 1077-1090.