THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF PSYCHOLOGY

Understanding Emotion in Different Languages: Emotional Prosody Recognition in Bilinguals and the Impact of Background Noise

GABRIELLE STOKES
Spring 2024

A thesis
submitted in partial fulfillment.
of the requirements
for baccalaureate degrees
in Majors Psychology and Science
with honors in Psychology

Reviewed and approved* by the following:

Michele Diaz
Professor of Psychology, Linguistics, and Neuroscience
Thesis Supervisor

Janet van Hell
Distinguished Professor of Psychology and Linguistics
Thesis Supervisor

Frank Hillary
Professor of Psychology
Honors Adviser

Funded by the NSF*
Electronic approvals are on file.

ABSTRACT

There are two main components to how we speak: what we say and how we say it. Currently, improvements in cross-cultural communication focus on translational efforts and how to get messages across semantically—what we say. However, among the most crucial components of connecting and communicating are the emotions we convey when we speak—the how. While extensive literature can be found on the concept of emotional prosody, very little is known about how acoustics are specifically linked to emotions and how environmental or cultural elements can influence recognition (Larrouy-Maestri et al., 2024). Previous studies have examined how humans recognize emotion in a foreign language that they don't speak, with an overarching goal to examine whether aspects of emotional prosody may have universal qualities (e.g., Paulman & Uskul, 2014; Pell et al., 2009). Their findings indicated that there is an "in-group advantage" in emotional prosody recognition where listeners more accurately recognize emotions in their native language compared to a foreign language (Pell et al., 2009). However, it remains to be seen how these effects apply to bilingual individuals. Our study investigated emotional prosody recognition between different languages in bilinguals, and investigated whether and if so, how this "in-group advantage" applies to individuals in their second language (L2) as well as their first language (L1). Dutch-English bilinguals listened to pseudo-sentences in Dutch (L1), English (L2), Arabic (foreign), and Hindi (foreign). These pseudo-sentences were created to match the phonetics and syntax of each language but contained meaningless words, allowing participants to only focus on the prosodic elements of the speech (an English example, "the fector egzullin the boshent"). For each language, participants listened to pseudo-utterances spoken with happy, sad, fearful, angry, and neutral intonation. After listening to each pseudo-utterance, participants determined which emotion they thought the speaker expressed, using a button press. To investigate the effect of background noise on emotion recognition (to mimic a real-world scenario), pseudo-utterances were presented in quiet or in two-talker Dutch babble. We found that Dutch-English bilinguals showed an advantage in correctly identifying emotional prosody in L1 and in L2 over the foreign languages Arabic and Hindi, both in the quiet condition and in the noise (two-talker babble) condition. These findings indicate that the in-group advantage not only emerges in bilinguals' first language but also in their second (and less proficient) language. We also found that participants had decreased emotion identification accuracy in the babble condition when compared to the quiet condition, and, surprisingly, that bilinguals were better at identifying emotions in L2 English than in L1 Dutch pseudo-sentences. Finally, earlier age of English acquisition and higher English proficiency (as measured by LexTALE; Lemhöfer & Broersma, 2012) were positively correlated with emotion identification accuracy. The results of this study will be related to current theories on emotional prosody and emotion recognition in the context of cross-cultural communication.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank everyone who helped me to complete this thesis.

I would like to thank Dr. Michele Diaz for first recruiting into her Language and Aging Lab two years ago. The skills I have learned and connections I fostered through that experience made me the researcher I am today. Dr. Diaz has always showed me support and granted me a multitude of opportunities to extend my learning in exciting new ways. I would not be here nor would my thesis be as it should be if it were not for the guidance Dr. Diaz has given me these past few years.

I would like to thank Dr. Janet van Hell for eagerly welcoming me into her Lab in my junior year in order to help me conduct this fascinating research. Dr. van Hell took me under her wing during the PIRE program and helped introduce me to the world of linguistics. She not only allotted me the incredible PIRE experience in which I conducted this research for her lab, she also helped greatly in the preparation of this thesis.

I would like to thank Adriana Miller for everything. Adriana and I traveled and conducted this research together. Following Adriana's guide, I was able to learn countless new research skills/programs that will be extremely beneficial to my career. Throughout this thesis writing process, Adriana has been there every step of the way—even with the preparation for research conferences. Not only that, after spending much time together, I not only appreciate her for her role as co-research assistant but also as mentor and friend.

I would like to thank Susanne Brouwer for being such a lovely host as we conducted our research at Radboud University. I truly learned so much in this experience abroad and had a fantastic time getting to know Dr. Brouwer and her intelligent, kind colleagues.

I would like to thank my family and friends for supporting me every step of the way. I had big dreams for my thesis and was eager to travel, so without my loved ones cheering me on and believing in me, I never would have been as successful. I am nothing without my peers, so having them to lean on during this large yet tasking accomplishment was essential.

This work is supported by NSF grant OISE 154900

Introduction

All forms of life interact through some form of communication. For humans, the main form is spoken language. There are two main components to how we speak: what we say and how we say it. Among the most crucial components of connecting and communicating are the emotions we convey when we speak—the how, and not just based specifically on the words said but also the emotions portrayed in the ways they're spoken. In fact, research has found that emotion can be identified without recognizable words at all, and all one needs are utterances comprised of pseudowords (Pell et al., 2009; Scherer et al., 2001). Instead, the prosody of language (such as tone, pitch, and rhythm) can be used to convey and also recognize emotion. In addition, not only can these prosodic elements help to portray emotion in one's own language, but studies show that these same effects come into play cross-culturally as well, because listeners could also recognize emotions when expressed in a language they do not speak (Pell et al., 2009). Nevertheless, there is an increase in the accuracy of emotional recognition when comparing listening to one's native versus a non-native language (Elfenbein & Ambady, 2002).

Multiple previous studies have examined how humans recognize emotion in a foreign language that they don't speak, with an overarching goal to examine whether aspects of prosody to express emotion in language may have universal qualities (e.g., Paulman & Uskul, 2014; Pell et al., 2019). These studies focused on monolingual speakers and used pseudo-utterances (e.g., "the fector egzullin the boshent") in order to eliminate any effects from the actual meaning of words. Pseudo-utterances, sentences formed by non-words that follow the grammatical and phonological patterns of a certain language, keep the participants from guessing the emotion based solely on the words themselves, and instead allows them to focus on the prosodic elements. Specifically, Pell et al. (2009) recruited 61 monolingual, Spanish speaking adults and had them listen and guess the emotion (joy, sadness, anger, fear, disgust) for pseudo sentences in English, German, and Arabic (languages foreign to the participants). Paulman and Uskul (2013) then replicated this study using both individuals familiar with the Chinese language (British being foreign) and individuals familiar with the British language (Chinese being foreign). Their participants also listened to pseudo utterances in each language but instead had to choose between seven different emotions (anger, disgust, fear, happy, neutral, sad, and surprise). Both studies found that participants were able to accurately detect emotion for all languages at a level greater than chance but had a greater accuracy when listening to their native language (Paulmann & Uskul, 2014; Pell et al., 2009; Scherer et al., 2001). This suggests that emotional prosody recognition contains both universal and cultural influences. Elfenbein and Ambady (2002) coined the term "in-group advantage" for this advantage in recognizing emotional prosody in one's native language. However, little research has tested whether and how this advantage extends to an individual's second language in bilingual speakers. Bilingual individuals represent a large part of our population yet are often underrepresented in research. They offer an interesting new perspective on this question, though, because bilingual brains are wired differently than monolinguals', and they are influenced by more than one culture, which affects their perspective on many matters, and possibly also emotion prosody recognition in language. By taking the bilingual aspect into account, we can further investigate the universality of emotion representation. If bilingual individuals can recognize emotion in their second language better than their foreign languages, this demonstrates that emotional prosody recognition can be learned, which in-turn can influence overall cross-culture communication.

Another factor that can affect the accuracy of emotional prosody recognition and overall speech perception is the presence of background noise. Laboratory-controlled experiments often lack the environmental factors of everyday life, such as busy streets or noisy restaurants, which impact language comprehension and intelligibility. By simulating noisy environments, researchers can get a clearer depiction of how accurate emotional prosody recognition is in more realistic circumstances. Previous studies have found that adverse conditions have an effect on speech perception due to multiple types of linguistic interferences (i.e. phonemic, acoustic, semantic, etc.), making it more difficult to recognize emotions when noise is present (Brown et al., 2022, Mattys et al., 2012). For bilingual individuals, their performance on speech perception accuracy has been found to be much more diminished for their second language than their first language when in “noisy” conditions (Tabri et al., 2010). However, the more the background noise phonetically matches the target stimulus, the more of an interference the background noise will have (Brouwer et al., 2012). Brouwer et al. (2012) investigated this by having Dutch-English bilinguals recognize speech in either Dutch or English with either Dutch or English background noise. They also set to test whether semantically significant background noise had more of an impact on speech recognition. Overall, they confirmed that the closer the background noise is to the target speech, the more interference it will have (with semantics considered as well). To illustrate, if the target speech is English pseudo sentences, then the background noise with the most interference will be that which is also English pseudo sentences. They also found that the more familiar one is with the background noise language, the more of an impact it will have on target-speech recognition (Brouwer et al., 2012). Therefore, we predict that the target-masker effect has a large impact on bilinguals who may have phonetic interferences from multiple languages. We will investigate the impact of background babble in one’s native language while trying to predict the emotional prosody of speech in their native, second, and foreign language(s). Based on Brouwer et al. (2012), we predict that background noise that matches the target speech (i.e. Dutch babble over top of Dutch speech) will have a negative impact, while the impact of the same, native babble on a target speech that is not the same language will be less. We believe this target-masker effect will also apply to emotional prosody recognition, so we hope to confirm and build upon these findings as we analyze the effects of the native-language babble condition.

Present Study

The present study will expand on Pell et al. (2009) by examining bilinguals rather than monolinguals, and will address the question whether, and how, bilinguals' emotional prosody recognition levels is different in their first and second language, relative to unknown languages. This question will be tested by having Dutch-English bilingual participants determine the emotion of pseudo-sentences spoken in Dutch (L1), English (L2), and two unknown languages (Hindi and Arabic). The basis of this study is to test whether the "in-group advantage" of emotional prosody recognition for the first language also extends to a second language. The second question will address the effect of unfavorable listening conditions on emotional prosody recognition across the different language types. Listeners will determine the emotion of pseudo-sentences for each language by selecting the emotion from a series of options (angry, fearful, sad, happy, neutral). This task will be performed with no background noise and also with multi-talker Dutch babble. The goal is to gain a better understanding of the accuracy of emotional prosody recognition in everyday life.

We predict that all emotions will be recognized at a level greater than chance across all languages, but that native language will have the highest accuracy out of all the languages. As for second language, we predict that emotion recognition accuracy will be higher than that for foreign languages but not as high as native language. Also, we hypothesize that introducing background noise will decrease emotion recognition accuracy overall. Because we are utilizing a

Dutch-two-talker-babble background noise condition, we predict this will have the greatest negative impact on the Dutch language condition then the second most on English because English has similar phonetic elements to Dutch. However, since our background condition is using semantically meaningless sentences with real words and our stimuli are pseudo-sentences, we do not predict this condition to have as large an impact as it would if it was also pseudo-sentences. We predict limited effect on the foreign languages from the background noise condition.

We ran our study to build upon the Pell et al.'s (2009) finding that one can recognize emotions for all languages at a level greater than chance and determine if Elfenbein and Ambady's (2009) "in-group hypothesis"—one is better at recognizing emotions in their native language than in a foreign language—also applies to bilingual speakers' second language. Also, we want to see if the presence of two-talker Dutch babble in the background will negatively impact one's emotional prosody recognition. Along with this, we also want to test Brouwer et al. (2012)'s target-masker similarity hypothesis to measure whether Dutch and the languages more similar to Dutch (English) are more negatively impacted by the Dutch babble than the less similar languages (Hindi, Arabic)

Methods

Participants

We recruited and tested 32 native Dutch speakers who also spoke English as their second language and had no knowledge of Hindi or Arabic. Of these 32 participants, we excluded data for two of them due to one participant having self-reported brain damage and another having a low outlying score in overall emotion recognition. The final sample included 30 participants with an average age of 25.13 years (*SD* = 9.05; minimum = 18; maximum = 60). The average score on the LexTALE English proficiency test was 86.79% (*SD* = 10.84, minimum = 66.25%; maximum = 100%) and the average age of English acquisition was 9.23 years (*SD* = 2.13). 17 participants identified as female, 10 as male, and 3 as non-binary/third gender.

Materials

Stimuli included English, Arabic, and Hindi pseudo-sentence recordings from Pell et al. (2009A). We created Dutch pseudo-sentence recordings with two native Dutch speakers following the procedures from Pell et al. (2009B). Each speaker was instructed to speak each of the 30 pseudo-sentences with prosody corresponding to each emotion condition: anger, fear, sadness, happiness, and neutral emotion. The last pronunciation of each sentence (unless there was a notable mistake) was selected and adjusted using Praat (Boersma & Weenink, 2024) for further analysis. After recording the pseudo-sentences, we conducted a norming study using Qualtrics. The norming study included 30 native Dutch speakers (separate from those in the experiment), recruited via Prolific, currently living in the Netherlands (21 male; 9 female; mean age = 28.53, *SD* = 9.94). Results from the Dutch norming study and previous English, Arabic, and Hindi norming studies conducted by Pell et al. (2009B) were used to determine the most highly rated speakers and sentences.

The experiment included 60 sentences per language (4) for each noise condition (2), totaling in 480 sentences. One speaker per language was chosen based on whichever was ranked more accurately by native speakers, and the top 12 sentences per emotion for each language were included.

To create the babble used in the noise condition, we recorded two native Dutch speakers producing various sentences in Dutch. We then took the recordings of the two speakers and overlapped their files in order to create the babble noise. We then applied the babble to the same sentences chosen for our quiet condition in order to reduce variability within the conditions.

Procedure

The experimental procedure followed that of Pell et al. (2009A). Specifically, participants were presented with one pseudo-sentence at a time, and the order of language blocks was counterbalanced. Within each language block, the emotion of each pseudo-sentence was pseudo-randomized so that no two sentences with the same emotion appeared back-to-back. Each language block began with the quiet condition followed by the noise condition. Each sentence was audibly presented via headphones one sentence at a time. Participants were asked to determine the emotion of the sentence using a mouse-click (presented with five choices labeled: happy, sad, fearful, angry, and neutral). Participants were not notified whether their response was correct or incorrect. Within each language block participants were allowed a break between the quiet and babble noise conditions and between each language block. After all trials were completed, participants then completed a series of tests and questionnaires to measure language proficiency and cognitive ability including: a debrief feedback survey about the experiment, a language history questionnaire, the LexTALE task to measure their fluency in English (Lemhöfer, & Broersma, 2012), and finally the Schutte Self-Report Emotional Intelligence Test (Schutte et al., 1998). After these questionnaires, the experiment was then concluded.

Analysis

Data were analyzed using R. A three-way ANOVA was conducted to determine how accuracy on the emotion recognition task was affected by language (English, Dutch, Arabic, Hindi), noise condition (quiet, noise), and emotion (anger, fear, sadness, happiness, neutral). Where applicable, post-hoc pairwise comparisons were conducted using Tukey post-hoc test.

Results

A 4 (language) x 5 (emotion) x 2 (noise condition) three-way ANOVA was conducted on the emotion recognition accuracy data, which showed a significant three-way interaction between the variables, $F(12, 14360) = 2.17$, $p = .01$. There were also significant two-way interactions between each of the variables: language and emotion, $F(12, 14360) = 89.23$, $p < .001$; language and noise condition, $F(3, 14360) = 2.70$, $p = .04$; and emotion and noise condition, $F(4, 14360) = 10.56$, $p < .001$. Mean accuracy by language, accuracy, and noise condition can be seen in Table 1 and are depicted in a box plot in Figure 1.

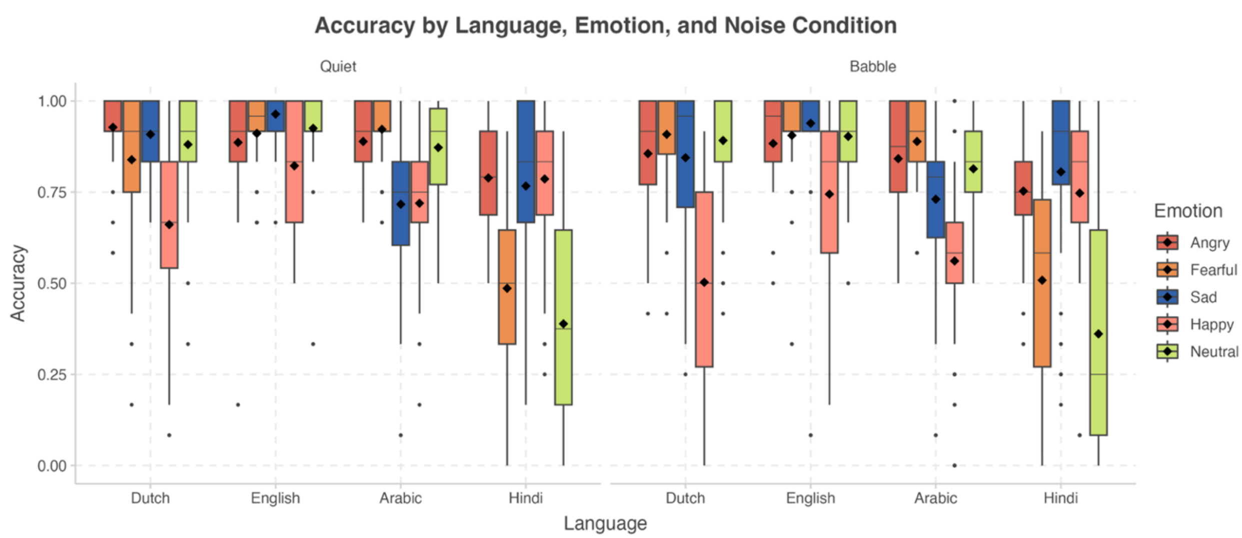

Table 1. Mean accuracy by language, emotion, and noise condition.
Figure 1. Box-plot of 3-way interaction between emotion recognition accuracy, language, and emotion between the two noise conditions (quiet and two-talker Dutch babble).

Noise Condition	Language			
	Dutch	English	Arabic	Hindi
		Quiet		

Anger	92.78%	88.61%	88.89%	78.89%
Fear	83.89%	91.11%	92.22%	48.61%
Happiness	66.11%	82.22%	71.94%	78.61%
Sadness	90.83%	96.39%	71.77%	76.67%
Neutral	88.06%	92.50%	87.22%	38.89%
		Two-Talker Dutch Babble		
Anger	85.56%	88.33%	84.17%	75.28%
Fear	90.83%	90.56%	88.89%	50.83%
Happiness	50.28%	74.44%	56.11%	74.72%
Sadness	84.44%	93.89%	73.01%	80.56%
Neutral	89.17%	90.28%	81.39%	36.11%

Emotion Accuracy Between Languages

Overall, when collapsing across noise conditions and emotions, accuracy was highest for the L1 and L2 languages compared to the foreign ones, as seen in Figure 2. To illustrate, Dutch accuracy was greater than Hindi ($p < .001$) and Dutch accuracy was greater than Arabic ($p = .02$). Finally, English accuracy was greater than Dutch (and therefore greater than the two foreign languages) ($p < .001$). This also concludes that L2 had higher emotion accuracy than L1 for the bilingual individuals.

However, when we look specifically at each language in each noise condition as seen in Figure 3, we find the same trend of Dutch accuracy being greater than Hindi ($p < .001$) and English accuracy being greater than Dutch ($p < .001$), yet we find that the Dutch accuracy is equivalent to Arabic accuracy in both quiet ($p = .83$) and noise conditions ($p = .19$). For both Arabic and Dutch, quiet condition accuracy was greater than babble condition accuracy (both $p < .001$), while accuracy was equivalent in both conditions for English ($p = .48$) and Hindi ($p = .99$). Overall, when looking solely at noise conditions, collapsing across both language and emotion, emotion recognition accuracy is much greater in quiet conditions compared to babble conditions ($p < .001$), as seen in Figure 4.

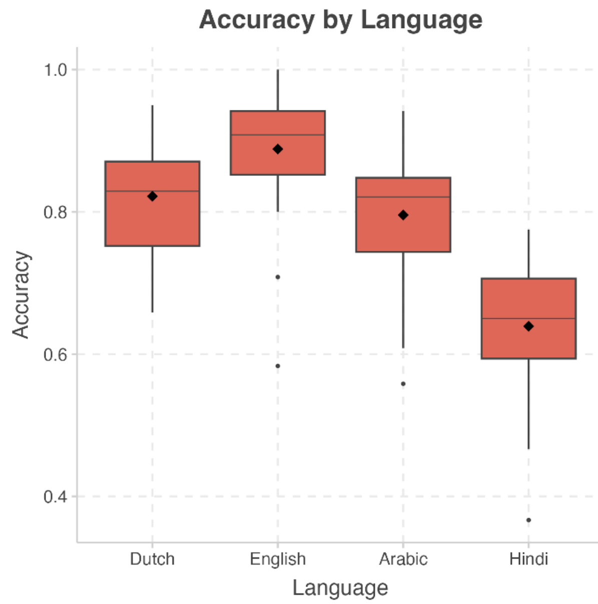

Figure 2. Boxplot of emotion recognition accuracy by language collapsed across noise conditions.

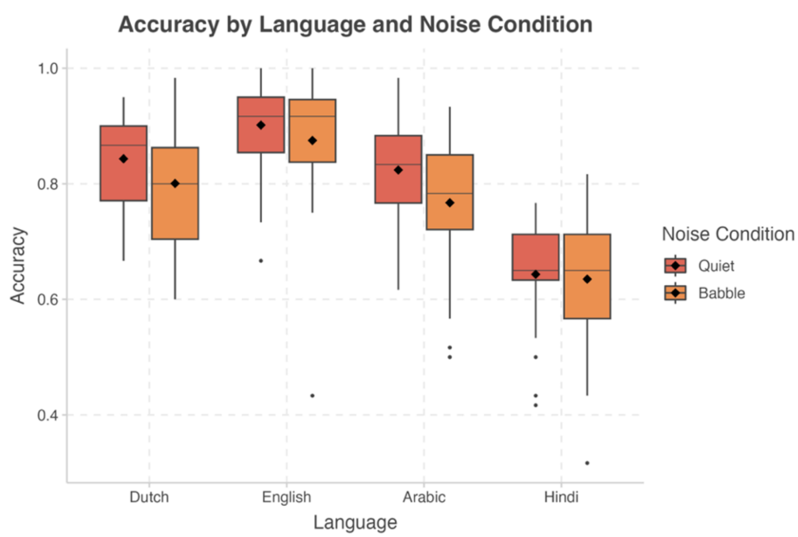

Figure 3. Boxplot of emotion recognition accuracy by language and noise conditions (quiet and two-talker Dutch babble).

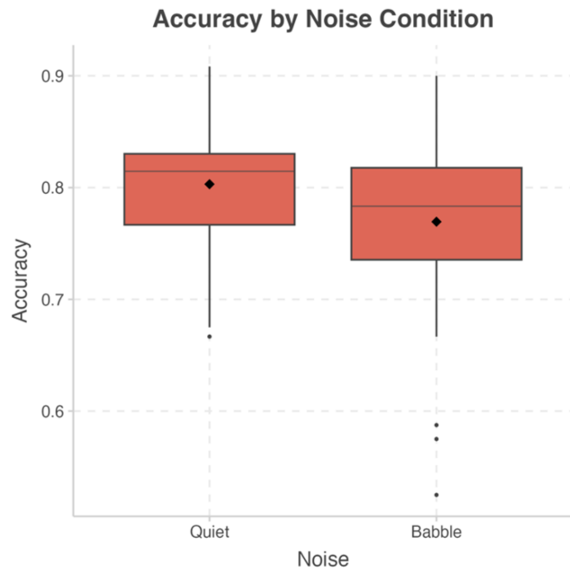

Figure 4. Boxplot of emotion recognition accuracy by noise condition (quiet and two-talker Dutch babble).

Age of English Acquisition and English Proficiency

Correlations were run to test the relationship between English emotion accuracy and language measures: age of English acquisition and English proficiency. English proficiency (measured by LexTALE score) was positively correlated with emotion accuracy (r = .48, p = .007) as seen in Figure 5. The average LexTALE score for participants was 86.79, which is a relatively high average. Age of English acquisition was also moderately correlated with emotion accuracy for (r = -.32, p = 0.079), as seen in Figure 6.

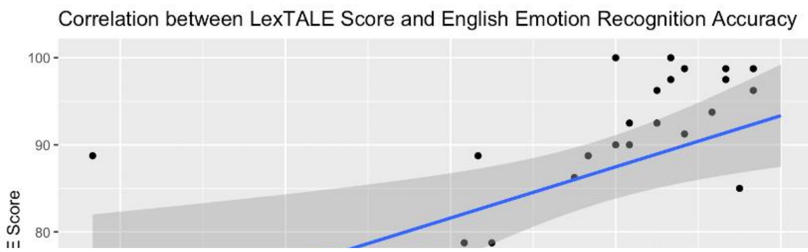

Figure 5. Scatterplot of average LexTALE Score by average emotion recognition accuracy in the English condition.

Discussion

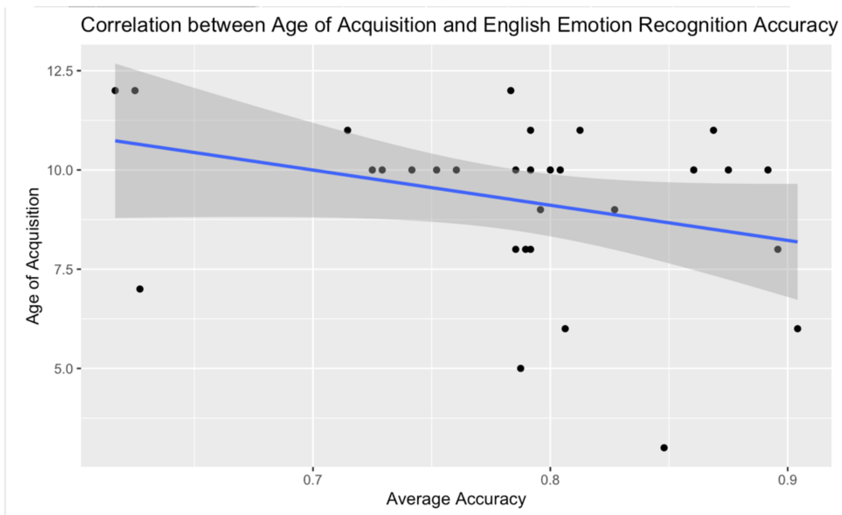

Figure 6. Scatterplot of average age of acquisition by average emotion recognition accuracy in the English condition.

Upon analyzing our data using a three-way ANOVA of language (4) x emotion (5) x noise condition (2), we found significant interactions between all variables. The main hypothesis of this study was to test to see if the “in-group advantage” found in previous research (including Pell et al., 2009A) was extendable to one’s second language. In the Pell et al. (2009A) study, it was found that all participants could accurately recognize emotion at a level greater than chance for all languages. However, they found that emotion recognition accuracy is greatly increased in one’s native language compared to a foreign language. Therefore, we wanted to see how this effect applied to bilingual individuals and their second language.

With our findings, we were able to confirm that bilinguals can recognize emotion in all languages at a level greater than chance. We also found that across noise conditions, emotion recognition accuracy was increased for first (Dutch) and second language (English) when compared to the foreign languages. However, we actually found that the accuracy for second language (English) was significantly higher than the recognition accuracy for first language (Dutch).

While this was unexpected because we predicted that first language would either be equal to or have greater accuracy than second language, we do have to take into account the prevalence of the second language. Our study was run in the Netherlands with Dutch-English bilingual participants with Dutch as their first language and English as their second language. As far as second languages go, though, English is highly prevalent in the Dutch community when compared to other cultures and their second languages. While we measured for English proficiency, we did not ask for amount of use, which may be the more deciding factor because the more often English is used (especially in contexts with other languages in the background), the more practice one has recognizing emotions regardless of their proficiency level. We must also consider the possibility that English may just overall be an easier language to recognize emotions in due to the expressive nature of the culture.

Also, we have to take into account the accuracy of our stimuli based on the results from norming studies with native speakers of each language. When comparing the rankings of our chosen sentences per language, English had the highest averages for their sentences, then Dutch, then Arabic, then Hindi, which directly mimicked our accuracy results, so their norming ranking may have been an influential factor in their final accuracy results. However, when we looked at emotion recognition accuracy in only quiet conditions, we found that English recognition accuracy was still greater than all others, but Dutch recognition accuracy was actually equal to that of Arabic (Hindi was still the worst). Again, this finding is very unexpected and goes against the Pell et al. in-group hypothesis but may also be explained by the differences in the accuracy data from the norming studies of our stimuli.

English accuracy was also found to be positively correlated with English proficiency, which was measured by the LexTALE test. Individuals who had a stronger grasp on the English language performed better on the English emotion recognition task, which was predicted because those who are more familiar with the language or know it better should be able to recognize it better in both quiet and adverse conditions. A majority of our participants were highly proficient in English which could explain as to why participants performed extraordinarily well/better in English. Further research should compare the correlation between LexTALE scores and English emotion recognition accuracy to the correlation between proficiency level in native (and foreign) language(s) and their subsequent emotion recognition accuracy. This comparison would allow researchers to see if emotion recognition has to do less with age of acquisition of a language (i.e. native vs second language) and more with proficiency in a language.

We also wanted to test to see how background noise affected emotion recognition accuracy so that our results could mimic real-world scenarios. As predicted, the presence of background noise decreased overall emotion recognition accuracy across languages but has specific effects on each language based in similarity. Our background noise-condition was a two-talker Dutch babble, which was two native Dutch speakers speaking actual sentences that were then overlayed on top of each other. According to the study done by Brouwer et al., (2012) the closer related the background noise is to the stimuli, the greater the inhibitory effects the background noise will have on emotion recognition. With that, we then predicted that our Dutch babble condition would have the greatest effect on the Dutch emotion recognition accuracy and then English because English has very similar linguistic elements to the Dutch language. Nevertheless, while our findings portrayed that the Dutch babble condition negatively impacted the Dutch emotional prosody recognition accuracy, this condition also had a significant negative impact on the Arabic emotional prosody recognition accuracy (with no effect on English or Hindi emotional prosody recognition accuracy). After further analysis and consideration, we understood that the Arabic language also has very similar phonetic elements to the Dutch language, specifically the pronunciation of the Arabic "Beta" is very similar to the Dutch pronunciation of the letter "v" (Alighiri et al., 2020). Therefore, it makes sense why Arabic accuracy was affected. However, we are unsure as to why the Dutch-babble condition has no effect on the English accuracy. Again, part of it may be due to the prevalence of English in the Dutch culture, so people already have lot of practice with distinguishing English sentences from Dutch background noise.

Overall, this study helped to confirm many hypotheses brought on by Pell et al. (2009A) and Brouwer et al. (2012). To illustrate, we validated the results that one can recognize emotional prosody at a level greater than chance for all emotions. However, when looking at solely the quiet conditions, we found that Elfenbein and Ambady's "in-group hypothesis" only actually applied to second language in this case where English emotion recognition accuracy was much higher than all other languages. Also, we found no significant difference between the Arabic and Dutch recognition accuracy in the quiet condition, so we disputed the prior findings that emotion recognition accuracy in one's native language is significantly better than a foreign language (in this case Hindi). Again, though, Dutch accuracy was significantly better than Hindi accuracy, so this may be a better representation of the "in-group hypothesis" with a foreign language because Arabic and Dutch do in fact have similar linguistic elements in their languages. We did in fact find that noise conditions do have an effect on emotion recognition accuracy overall, but we also investigated Brouwer et al.'s (2012) target masker hypothesis where background noise that is more similar to the target speech will have greater inhibitory effects on emotional prosody recognition than less similar target speech. Therefore, we predicted the two-talker Dutch babble to greatly affect the Dutch and English (very linguistically similar to Dutch) emotional prosody recognition accuracies, yet the babble noise condition in actuality had effects on the Dutch and Arabic emotional prosody recognition accuracy. Upon researching further, we discovered the similarities between the Dutch and Arabic languages that may explain the task-masker hypothesis for the Arabic language condition. Nevertheless, it is still being questioned as to why the two-talker Dutch babble had no effect on the English language condition; it may be due to the prevalence of English in the Dutch culture. As a second language, English is much more commonly used than most other second languages, especially in the Dutch culture, so it may just be that participants are used to focusing on English sentences with Dutch babble in the background in real-world scenarios.

Future Directions

While we were able to confirm many prior findings and added new insights on bilingual listeners' recognition of emotional prosody in their known language relative to unknown languages, a lot of questions were also introduced as well. For example, we are unsure as to why English emotion recognition accuracy was much greater than all the rest, especially Dutch and why Dutch emotion recognition accuracy was not significantly different from Arabic emotion recognition accuracy. In order to better understand this, a follow-up study should be performed where participants must measure their proficiency in all languages, so we can then see how these values correlate to accuracy rather than just a statement of proficiency (Dutch) or lack thereof (Hindi and Arabic). Another potential confounding variable could be the accuracy of our stimuli. All sentences for all languages were submitted to a norming study with native speakers of each language, and the most accurate sentences for each language was chosen. However, there were distinct differences between the highest accuracies for some languages (i.e. English had exceptionally high accuracies) compared to others (Hindi had exceptionally low accuracies). If this study were to be repeated, it is suggested that stimuli are chosen based on most similar accuracies between languages rather than highest accuracies within language. Finally, to look further into the lack of effect the Dutch-babble had on the English language condition, we may need more anecdotal evidence or observational data to show how prevalent the English language truly is in their culture and if this simulated condition we made actually mimicked a very common situation for participants. This summer, there will be a follow-up study performed at the same university instead measuring a two-talker English noise condition to see how this will affect emotional prosody recognition in all languages. Future research should also look into running this study within other bilingual cultures that have other native/second language combinations besides Dutch and English to truly see if this is a universal phenomenon.

Through our combined work with that of Pell et al., (2009A), we were able to emphasize the fact that all individuals can recognize the emotion of pseudo-utterances at a level greater than chance for all languages. These results illustrate that there must be a universal component to how humans express emotion through the prosodic elements of speech otherwise people would have only recognized emotion accurately for languages they are familiar with. However, recognition was significantly better (in collapsed conditions) for English and Dutch languages (the familiar languages), implying that there must also be a cultural impact on the prosodic qualities of emotion expression. Nevertheless, the fact that Arabic and English emotion recognition were not statistically different in the quiet condition may provide evidence that prosodic emotion recognition has less to do with the similarities in the culture and more in the similarities of the languages specifically (phonetic elements/pronunciation). Also working with Brouwer et al., (2012)'s task-masker hypothesis, we found that the presence of background noise has an aversive effect on emotion recognition overall and specifically affects the accuracy for the languages that are most similar to the background noise. Again, though, in our case we found English to be unaffected by the Dutch-babble background noise, so we must consider other external factors as to why this might be the case (i.e. the role of English as a second language). Our results allow us to take a better look at the underworking’s of cross-culture communication because now we understand that there are both universal and cultural factors we must consider when trying to communicate emotions. The background noise also allowed us to see how emotional prosody recognition works in scenarios that mimic the real-world to see how these cultural and universal elements come together in real conversations.

BIBLIOGRAPHY

Alighieri, C., D'haeseleer, E., Daelman, J., Van Lancker, F., Laperre, M., Kissel, I., & Van Lierde, K. (2020). Articulation skills in bilingual children with a migration background: a comparison between bilingual Turkish-Dutch, Arabic-Dutch and monolingual Dutch children. *Journal of Communication Disorders*, *87*, 105993. https://doi.org/10.1016/j.jcomdis.2020.105993

Boersma, P. & Weenink, D. (2024): Praat: doing phonetics by computer [Computer program].

Brouwer, S., Van Engen, K. J., Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: language familiarity and semantic content. *The Journal of the Acoustical Society of America*, *131*(2), 1449–1464. https://doi.org/10.1121/1.3675943

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Larrouy-Maestri, P., Poeppel, D., & Pell, M. D. (2024). The sound of emotional prosody: Nearly 3 decades of research and future directions. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 17456916231217722. Advance online publication. https://doi.org/10.1177/17456916231217722

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

Paulmann, S., & Uskul, A. K. (2014). Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition and Emotion*, 28(2), 230-244. https://doi.org/10.1080/02699931.2013.812033

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009A). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, 33(2), 107–120. https://doi.org/10.1007/s10919-008-0065-7

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009B). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, 37,417-435. https://doi.org/10.1016/j.wocn.2009.07.005

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Domheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, 25(2), 167-177. https://doi.org/10.1016/S0191-8869(98)00001-4

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

https://doi.org/10.1177/0022022101032001009

Tabri, D., Abou Chacra, K. M., & Pring, T. (2011). Speech perception in noise by monolingual, bilingual and trilingual listeners. *International journal of language & communication disorders, 46*(4), 411–422. https://doi.org/10.3109/13682822.2010.519372