2022

FACE MASKS AND SPEECH PERCEPTION: EMOTIONS AND INTELLIGIBILITY PERCEIVED BY MONOLINGUAL AND BILINGUAL SPEAKERS

Monica Andrea Chulde Guayasamin
University of Kentucky, moni.chulde@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0001-9376-3266
Digital Object Identifier: https://doi.org/10.13023/etd.2022.004

Monica Andrea Chulde Guayasamin, Student

Dr. Kevin B. McGowan, Major Professor

Dr. Allison Burkette, Director of Graduate Studies

FACE MASKS AND SPEECH PERCEPTION:
EMOTIONS AND INTELLIGIBILITY PERCEIVED BY MONOLINGUAL AND BILINGUAL SPEAKERS

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Arts in the College of Arts and Sciences
at the University of Kentucky

By

Monica Andrea Chulde Guayasamin

Lexington, Kentucky

Director: Dr. Kevin B. McGowan, Professor of Linguistics

Lexington, Kentucky

2021

https://orcid.org/0000-0001-9376-3266

ABSTRACT OF THESIS

FACE MASKS AND SPEECH PERCEPTION: EMOTIONS AND INTELLIGIBILITY PERCEIVED BY MONOLINGUAL AND BILINGUAL SPEAKERS

Speech perception in unfavorable conditions reduces the intelligibility of the message. The use of face masks may be one factor that degrades the comprehension of target words in transcription tasks and the recognition of emotional prosodies. Different researchers have proposed the influence of visual stimuli in the comprehension of the linguistic message (e.g., Tuomainen et al., 2005; Schwartz et al., 2004; Llamas et al., 2008; McGowan, 2015). This study reports the results of an experiment that tests how intelligibility and emotional prosody are affected by surgical masks. The online experiment has been applied to two groups of speakers from the University of Kentucky. The first group consisted of 42 students from the Latin Students Association (LSA) with Spanish-English backgrounds who were subdivided in three groups according to their language's history, monolinguals (L1), early bilinguals (L1-2), and late bilinguals (L2). The second group consisted of 10 monolingual English speakers from the Linguistics department. Both groups performed a transcription task plus an emotional prosody rating from 60 videos of a native English speaker in two conditions: MASK and NO MASK. Participants in both groups obtained more accurate results in the NO MASK perception tasks.

KEYWORDS: Speech perception, surgical masks, intelligibility, emotional prosody, language history

Monica Andrea Chulde Guayasamin

December 08, 2021

FACE MASKS AND SPEECH PERCEPTION:
EMOTIONS AND INTELLIGIBILITY PERCEIVED BY MONOLINGUAL AND BILINGUAL SPEAKERS

By
Monica Andrea Chulde Guayasamin

Dr. Kevin B. McGowan
Director of Thesis

Dr. Allison Burkette
Director of Graduate Studies

December 08, 2021
Date

DEDICATION

To my family, those who are present now and all those who brought me here.

ACKNOWLEDGMENTS

I am thankful with all my professors in and out the Linguistics department. Thank you to the following thesis committee members, my Thesis Chair, Dr. Kevin McGowan, who provided timely and instructive comments and evaluation at every stage of the thesis process. Next, I wish to thank the complete Thesis Committee: Dr. Mark Lauersdorf, and Dr. Rusty Barret who provided insights that guided and challenged my thinking, substantially improving the finished product.

In addition to the technical and instrumental assistance above, I received equally important emotional assistance from family and friends who helped endure a global pandemic away from home. My fellow students, Kyler Laycock, Nour Kayali, Xueying Wang, Ruth Gonzalez, provided on-going support throughout the thesis process, as well as technical assistance critical for completing this project. Thanks to Jason Car, who motivated me to pursue a master's degree at UK. Thanks to my cohort Crissandra, Xueying, Mohammad, Joey, Yasha who were always admiring each other progress. Finally, I wish to thank the respondents of my study (who remain anonymous for confidentiality purposes).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

Different situations have put people in unfavorable communication conditions. One of these adverse circumstances is the communication in a pandemic setting that might have become more challenging due to the extra barriers implemented that have obligated people to social distance and wear a face covering to cover their mouths and noses. People have taken these actions to protect themselves and avoid the spread of the virus, COVID-19, that since 2020 has continued to be a threat. Understanding a linguistic signal from a half-covered face in public might be more difficult especially for non-native speakers listening to native speakers. The different kinds of face masks might affect acoustics; however, it has been demonstrated that most of them "had little effect below 1kHz" (Correy et al., 2020). The surgical mask has been used for this experiment which, according to Toscano and Toscano (2021), has produced the smallest acoustic effect and has yielded the best performance in speech perception tasks.

To better understand the results from bilingual speakers in speech perception tasks, it is important to classify participants according to their different language backgrounds. Mayo et al. (1997) have organized four groups of language learners (Spanish English) by their time of exposure and history of L2 into monolingual speakers, bilinguals since infancy, bilinguals since toddler, and bilinguals post puberty. Similar classifications include monolinguals, high proficiency early bilinguals, and late bilinguals (Regalado et al., 2019). This classification has contributed to understanding differences in speech perception outcomes. Besides, the two auditory-visual conditions (MASK and NO MASK) presented to participants have played an important role to determine accuracy in the transcription task and emotional prosody rating. Previous findings (e.g., Tuomainen et al., 2005; Schwartz et al., 2004; Llamas et al., 2008; McGowan, 2015) point out that the visual information affects the perception of the linguistic information. This thesis aims to contribute evidence about how the lack of visual information, in this case by the use of a surgical mask, affects the accuracy in transcription tasks and emotional prosody ratings in monolingual and bilingual (Spanish-English) speakers.

1.1 Internal and external factors

Depending on the circumstances and the different backgrounds of people with languages, communication may encounter difficulties when it comes to clear perception. Some possible common elements that degrade linguistic signals include intrinsic and extrinsic factors. Intrinsic factors refer to the internal characteristics that every individual possesses and the person’s ability to comprehend the linguistic message. The listener's intrinsic factors can include age, hearing disorders, other health problems, communication skills, age of second language acquisition, emotional state, among others. Participants in this investigation have not presented hearing difficulties that could greatly affect the results. For this project, proficient English speakers have been tested in the tasks that include monolinguals, early, and late bilinguals. Bilingual participants were undergraduate and graduate students from the University of Kentucky who belong to the Latin Students Association (LSA). These proficient English speakers are students who speak English as a second language and have been admitted to the University of Kentucky because of their ability to communicate in English. The University of Kentucky admits students who prove their English proficiency by presenting TOEFL scores of 71 (IBT), 197 (cbt) or 527 (paper-based) or a minimum score of 6.0 on the IELTS (Office of Students Financial Aid and Scholarships, 2021).

The extrinsic factors that might affect perception include background noises, such as cars passing, other people talking, birds singing, air conditioners, or people using silverware, just to mention a few. These noises could affect the listener's ability to perceive the linguistic signal. The background noises are part of an every-day situation, and everybody has experienced a conversation where more sounds or noises have appeared around or just passed through. The ability to focus on one person's speech "when others are speaking at the same time" has been described by Cherry (1953) as the "cocktail party effect." Additionally, Cherry claims that the brain has a logical principle of probability-ranking which enables it to combat noise or disturbances. When background noises interfere with a normal face-to-face conversation, some secondary cues become primary cues that will improve the reception of the message. Parikh and Loizou (2005) showed that the vowel formant frequencies (F1, F2) were more reliably detected in noise conditions than in quiet. These findings show the people's ability to comprehend a message even in unfavorable listening conditions.

To make this experiment more challenging, an English multi-talker babble noise (Van Engen & Bradlow, 2007) was added to the input that along with the surgical mask might resemble some everyday encounters people have. The surgical mask was used in 50% of the input, the two conditions were MASK and NO MASK, while the multi-talker babble was used in both conditions. Previous findings show that the perception of speech is different when background noise is added to the input than when in quiet conditions. However, in a recent study, Toscano and Toscano (2021), who used different levels of background noise in the input, have shown that there is no significant difference in speech perception results specifically with the surgical mask; however, the results vary according to the levels of noise. A multi-talker babble noise can contain linguistic and non-linguistic noise. The linguistic noise can be parallel to the primary signal, or it can be a different language. According to Van Engen (2010) and Kilman et al. (2014), the parallel language used for the target input and the multi-talker babble noise has shown a competing and distracting effect when perceiving the speech signals.

The noise level of the background babble is another important point that can modify the participants' output. For this study, the multi-talker babble noise was added with +3dB loudness which was 3 dB quieter than the target utterances. The noise level of the multi-talker babble applied to the input has allowed the participants to perform relatively well in this experiment. Overall, extrinsic and intrinsic factors present at the moment of communication or just at the moment of perceiving a message can be distractors that can influence the results of a good and clear communication and understanding.

1.2 Positive and negative factors in speech comprehension

Different investigators have listed key factors that lead to better language comprehension. Some positive factors are listening to familiar voices (Nygaard et al., 1994), listening to familiar dialects, listening to more typical standard voices, listening to more frequent words (King & Sumner, 2014 p. 2914), using the same language variety, using simple syntax, and more. Comprehending speech from relatives, friends, and people who live around us, and in our community can be easier and fast. People from the same social or economic statuses or geographic communities could share similar interests, backgrounds, and resources, but most importantly for this study, they might share the same language variety that lead them to a better intelligibility of the linguistic message. Sharing the same language becomes an asset for good communication; however, there are more factors that can help. It is also important to emphasize that listeners can quickly adapt to unfamiliar speech (e.g., Baese-Berk et al., 2013).

For native and non-native speakers of a language, a fundamental part of intelligibility in communication depends on the social information conveyed through the speech, this means the extra and essential cues that are part of the speech signal, such as emotions, gender, age, attractiveness, health, etc. Some of these elements can be perceived between two or more people even before the words start, and sometimes at first sight. Through body language, a person could perceive if the person is happy, sad, or angry. At first sight, most people can see if they are going to speak to a young or an old person, or to someone they perceive to be a man or woman. Our eyes perceive an impression of the person, and we can even feel attracted to this person. All the human senses are important; however, some people can rate their importance differently. In a survey developed by Enoch et al. (2019) in the United Kingdom displayed that the people's most valued sense was the sight followed by hearing. Some people can greatly rely on the eyes as the first receptor of information in a face-to-face encounter or when the encounter only involves one party, for example, someone rating or describing a photo or video.

Besides these important aspects, having a clear and free speech delivery can be key in communication, it might mean not covering part of your face with any garment. In the year 2020, and for a long period, people were asked to wear face masks in public and to keep social distance to minimize the spread of COVID-19 and its variants. Some of these face masks can be potential influencing factors that might affect communication. The facial garments block the mouth and part of the nose, and the propagation of the sound is hindered, interfering with normal production and perception of the speech. Besides, depending on the distance between the speaker and the listener, speech recognition might also have different effects (Toscano & Toscano, 2021). Some researchers, for example, Fecher (2014) stated that the speech production of the speaker is absorbed by the facewear material. Saigusa (2017) considers that the facewear could affect how the talker produces speech, how far the sound physically propagates, and how the voice quality is affected. These findings could contribute to the idea of face masks interfering with a correct perception of the linguistic message.

However, Llamas et al. (2008) showed that the loss of speech intelligibility derives from the reduction of the speaker's visual information and not from the facewear and their fabrics. Fecher (2014), in her forensic speech study, also points out that the observer or listener’s judgement is compromised when they are presented with auditory and limited visual input. Likewise, Tuomainen et al. (2005), and Schwartz et al. (2004) described the contribution of audio-visual speech input to better hearing and understanding in contrast to audio-only conditions with background noise. In a face-to-face conversation, the eye and the ear play a key role in perceiving speech. Some researchers have suggested that the multisensory domain is present when the brain processes visual and auditory stimuli at the same time. A clear example of a multisensorial work is the McGurk effect (McGurk & MacDonald, 1976) which demonstrated that the visual information does affect the perception of speech. Another example of how visual information affects the output that is also related to social categories is a study from McGowan (2015). In his work, McGowan recruited Chinese listeners for transcription tasks while they were presented with Chinese audio plus Chinese faces, Caucasian faces, and a person’s silhouette; these stimuli also contained background noise. The results showed that transcribers performed better when seeing the parallel between Chinese faces and speech.

In a face-to-face conversation, both sides will go back and forth interchanging thoughts and reading each other's social cues immersed in movements, tone of voice, and in the person per se. Conversely, in speech perception experiments, the listener will not receive feedback from the talker. However, the listener will be able to read cues from the talker (for this work, the participants have been asked to perceive emotions from the input). For this online study, participants have been recommended to use headphone/earphones to try to avoid distractions from the environment, such as air conditioners, music, and other external noises. Nevertheless, students who volunteered to participate in this study could do it by themselves in the place of their choice, using different devices such as desktops, laptops, cellphones, or tables.

The participant's predisposition and motivation are some factors that might affect the results as well. Different motivations as incentives can also make the participant improve their attendance to this kind of studies (e.g., offering extra credit in one assignment from a class). However, the LSA group was not offered any extra credit or any other motivation to complete the survey. For this study, 42 replies to the survey were obtained from the LSA group. Though in Johnson and Newport's (1989) study, motivation was a variable that did not correlate significantly with the performance of tasks (p.83). Being an online study, where no one is checking the participants' progress, time spent, or degree of focus on the experiment could have intervened as an important factor that has led to the currently outcomes. Respondents might have been distracted while completing this experiment. In different circumstances, this experiment should have been performed in person, in a lab with the appropriate equipment that might have improved the attendance and participation of respondents.

1.3 Face masks

Now, with the constant use of medical masks, more studies have demonstrated that all different masks alter the frequency and the directivity of speech in different levels. Correy et al. (2020) tested twelve types of face masks and concluded that most of them "had little effect below 1kHz, but they attenuated higher frequencies by different amounts. The surgical mask and KN95 respirator had peak attenuation of around 4dB." Likewise, Goldin et al. (2020) measured how speech was distorted using face masks, and they affirmed that the surgical mask acts as a low-pass filter that reduces the high frequencies from 3 to 4 dB. Lately, experiments with surgical masks have shown that the acoustics are not significantly affected. Toscano and Toscano (2021) also showed that face masks affect the speech signal by attenuating the frequencies and the directivity of the signal. Some researchers concluded that the surgical masks have produced the smallest acoustic effect obtaining the best performance (Toscano & Toscano, 2021). For this project, a native English speaker was recorded in two conditions, with MASK and NO mask. For the MASK condition the speaker wore a blue surgical mask. Figure 1 shows a surgical mask, which is made of polypropylene, has three layers, and is 0.4 mm thick.

Figure 1.3.1

Surgical Face Mask

Note. Face mask used by the speaker to record the MASK condition.

Different kinds of masks, plastic shields, or veils, for example, can be considered barriers to a normal two-way communication that might affect the correct comprehension of the linguistic signal. For instance, there have been some cases of communication problems because of veils. Two controversial incidents drew attention of people when different Muslim women wore a veil at a doctor's office. These "specialists" lived and worked in western places, different from the women's religion and culture. Jones (2019) in *The Daily Mail* from the United Kingdom cites cases of Muslim women who wore their veils and were asked to remove them for better communication. In the first case, a doctor asked the woman to remove her veil because he felt uncomfortable talking face to face to someone he could not see. The second case also occurred at a doctor's office, where the doctor asked the woman to remove her veil because her voice was muffled, and it was difficult to understand her. On one hand, these cases might show us that society still needs to learn to be respectful with other cultures, religions, and ways of living. Now, due to pandemic conditions, we had to learn and accommodate ourselves to live and communicate with facial garments in public. On the other hand, we need to think that we all are different, and some people might have the need to retrieve more information about the other person with more visual information. When a new person with a different language or culture communicates with us, there might be some constraints for any of the parts. Some people might need more details to comprehend the message, such as the visual information (e.g., seeing the lips movements), especially when the language is not the same.

We can find different kinds of face masks with different fabrics, colors, layers, and designs that people wear in public now. In the face covering group, we can also include items, such as glasses, sunglasses, helmets, ski-masks, and other garments that are normally allowed in public and are part of daily life in our society. Likewise, in the past, some people in European countries wore face garments with different purposes, for example half-masks for Carnival in Venice or black velvet half-masks in France as part of their culture. Many ancient cultures utilized face coverings that related to religious purposes, such as Islam. In ancient Mesopotamia and Persian empires, the veils also represented higher economic status. Societies, where the veil is worn by different people are accustomed to this garment and have not reported any lack of intelligibility in communication. According to Borneman (2009) some simple explanations of why to use a veil in certain cultures include "ecological reasons, protection from the desert sands, and symbolic, to ward off evil in encounters with strangers." Additionally, Borneman includes some women's reflections about the use of the veil, describing their wish to escape males' visual objectification and showing anonymity and defense against sexual harassment.

Depending on the society or culture and according to the situation, some masks or face garments have been associated with disguise, duplicity, sexual license, and crime, specifically in western countries (Winet, 2012). Society, its weather, different religions, or in this case a contagious disease can dictate what needs to be done or worn in every social group during a period of time. The social norms and behaviors are constructed through time. As described above, other garments that can cover part of the talker´s face can be helmets, scarves, hats, or even piercings. Some of these accessories are allowed in certain societies and banned in others. Nowadays, surgical masks and other respirators have become part of our daily routine when going out in public; and now wearing a face covering could represent respect to others. Regardless of the reason, every piece of facial garment described above can hide a part of the speaker's face that might make us lose some socio-indexical details and emotional prosodies when communicating with each other.

1.4 Socio-indexical properties of speech

As Ochs (1992) states "indexicality is represented as a property of speech through which cultural contexts, such as social identities and social activities are constituted by particular stances and acts." Socio-indexical cues include facial expressions, voice quality, rhythm, and more elements that form our perception of the person or speaker we are communicating with. All these cues and some others are retained in our memory, and we can use them to compare future encounters with different people and to find similarities. Docherty and Foulkes (2014) consider that the associations a person makes when talking with someone else are stored in their memory and are updated with each new encounter or new experience (p. 46). Also, Johnstone and Bean (1997) state that every person has different linguistic memories, and there are not two identical speakers. People have memories but not every memory is the same, and how the linguistic message is understood from different people we communicate with can be explained by the normalization assumption. This hypothesis assumes that even when people possess different phonetic apparatuses, people can understand each other because of the lexical-semantic content of speech (Jackson & Morton, 1984). Moreover, Goldinger (1996) supports that some voice details are discarded after the concept is retrieved, and it is stored in the long-term memory. Additionally, the listeners and their experiences, language background, expectations, social biases, cognitive abilities, and motivations are key factors for efficient communication between native and non-native speakers (Baese-Berk et al., 2020).

It has been agreed that the socio-indexical properties can include gender, age, ethnicity, or class (e.g., Johnstone & Bean, 1997). The socio-indexical cues, emotional prosody, and linguistic information are intertwined within the speech signal and the listener might understand the utterances and much more. Additional information a listener perceives is taken from different elements that the speaker shows in the speech, such as hesitation, approval, ordering, etc. The social information travels through the linguistic signal that includes phonology, morphology, syntax, pragmatics, semantics, and lexicon. Perceiving a linguistic signal includes more than words and can help us recognize and understand more about the speaker and the situation.

In a regular two-way communication, both parties receive the lexical-semantic message, plus other social meanings. The lexical-semantic production refers to a word, a phrase, a clause, or any linguistic structure. Some social indexes can be more obvious than others, one of them is gender. According to Ochs (1992), gender ideologies are socialized and sustained through talk. Besides, gender identity and other social signals may also depend on social constructs, behaviors, and their relationship with the language (p.337). For instance, every community or society assigns certain norms and linguistic forms to distinguish men and women's speech. The members of the same community can better interpret the linguistic message and the socio-indexical properties in an unconscious way due to everyday interaction. The message sent will always vary and will depend on the context in which it is said, the difference in social class, the familiarity, the relation between the speaker and addressee, and the emotional state of any of the parties just to mention some.

1.5 Body language and emotions

Likewise, a key part of communication is body language. Part of body language studies focuses on the facial gestures someone produces. Ekman and Friesen (1986) report seven universal microexpressions that people articulate; they are fear, surprise, sadness, happiness, anger, disgust, and contempt. Vanessa Van Edwards on the website *Scienceofpeople* (2020), explains that a microexpression is “a brief involuntary facial expression that appears on a person's face according to the emotions being experienced. Unlike regular, prolonged facial expressions, it is difficult to fake a microexpression.” According to this online scientific portal, people can recognize these microexpressions even if the person is wearing a face mask. A person can read facial expressions from the eyes up, this can be enough to retrieve some emotional prosody from a person. However, depending on the culture or society, the gestures for communication can vary. For instance, in Asian cultures engaging eye contact can be considered rude; conversely, in western countries it might mean something different.

Several researchers have worked with emotional prosody (affect) perception and priming; for instance, a study by Kim and Sumner (2017) where non-emotional words were used to test whether emotional information activates words associated with the corresponding emotion demonstrated an affecting priming result of the corresponding emotion. Kim and Sumner evidenced that an angry or happy prosody enables the recognition of words in two ways, the emotional information is conveyed independent of the word, and it also activates semantically associated words. Additionally, the researchers showed that some participants perceived more neutral emotions as happy and associated it to happy words. In another study, Wurm et al. (2001) stated that the emotional prosody can often be distinguished even in the absence of a speech signal, only by body language or the person's behavior. Various studies have proven that the listeners' comprehension of the linguistic message improves significantly when the message is accompanied by gestures. The facial speech shown to the listener facilitates the intelligibility of the spoken message especially in noisy conditions (Sueyoshi & Hardison, 2005). All these previous studies agreed on the importance and effect of visual stimuli on speech perception tasks.

1.6 Monolingual and bilingual speakers

When the interlocutors communicate using the same language that both have as L1 might increase the intelligibility of the linguistic signal. However, there might be cases where different varieties of the same language, or regional varieties, can interfere with a clear communication, besides age, status, gender, and others. Language varieties can vary in speed, meaning, or pronunciation. Being more specific, every person articulates differently, and everyone possesses distinct articulators that make speech production unique. Some of these differences could cause a misinterpretation of the message. If a misunderstanding happens among native speakers of the same language, it might be more probable that more misunderstandings occur among native and non-native speakers of a language, in this case, English and Spanish. As mentioned before, Sueyoshi and Hardison (2005) point out that for second language learners the auditory-visual input is important to interpret the nonverbal messages. It is also important to clarify that participants in general and specifically bilingual speakers presented with visual stimuli, would not behave as in normal conditions when they perform language experiments (Leather, 1999).

For this study, it is relevant to identify the age of second language acquisition of respondents to better understand the results. Mayo et al. (1997), in their study, worked with different participants with different language background who were divided according to age of exposure to English. The group was subdivided into bilingual since infancy (BSI), those who grew up with a Spanish speaker parent and the other parent as an English speaker. Bilingual since a toddler (BST), bilingual children who learned English as a second language before the age of 6. And bilinguals post puberty (BPP), speakers who learn English after 14 years old (p. 687). This classification was organized according to the pool obtained by the researchers. Overall, performance of BPP shows lower scores compared to the other groups. Similarly, in a masked speech perception experiment, Regalado et al., (2019) have grouped participants as follows: monolingual listeners, high proficient early bilingual listeners, and late bilingual listeners. The outcomes obtained by Regalado et al. show that late bilinguals performed more poorly than the other two groups.

Tabri et al. (2011) have recognized some elements that decrease the degree of performance in communication for non-native speakers, some of them are: the person's amount of exposure to the L2 language, age of L2 acquisition, and the poor listening conditions in the environment. Bilingual speakers might have a strong advantage when their exposure to the L2 occurred at an earlier age. Many researchers have studied this L1 and L2 early acquisition time known as the critical period. For instance, Lenneberg (1968) considered that a language could be acquired only in a limited period, from the early infancy until puberty. Later, Johnson and Newport (1989) expanded this critical period hypothesis to L2 speakers explaining that the earlier the speaker learned the L2, the better they would perform. The results of Johnson and Newport' investigation not only showed the performance in grammatical structure of English but also in self-consciousness, and homeland identification. Their results indicated the higher they performed, the more they identified with the country.

It can be difficult to understand how bilingual speakers process the L2. The bilingual language processing involves the entire linguistic system that would take another whole chapter to explain. However, some theories can be mentioned of how lexicon and semantics might work in a bilingual brain. Potter et al. (1984) suggested different hypotheses that might work with L2 learners, that might also depend on their L2 fluency. The researchers cited the *word association hypothesis* which works when the L2 remains weaker. This possible hypothesis considers a direct association of an L2 lexical item with the corresponding lexical item in L1. Another premise presented by Potter et al. is the *concept mediation hypothesis* which is associated directly with the underlying concept. Eventually, the word association process might be replaced by the concept mediation stage (p.24). There are several L2 theories that propose how bilingual speakers' brains work to process language. At the end, the results of social speech perception experiments with monolingual or bilingual speakers might be seen and analyzed as a derivation from respondents' ideologies and biases (Kang & Rubin, 2009).

CHAPTER 2. METHODOLOGY

This study was approved by the University of Kentucky Institutional Review Board (IRB). The main experiment consisted of 60 clips that showed a native English speaker uttering individual sentences. The talker was a 24-year-old male born in Western Kentucky. He was also a master's student from the University of Kentucky. The speaker fitted the surgical mask how he usually wears it in everyday settings, and in a way he felt comfortable with it. The blue polypropylene (YY/T 0969) surgical mask was used for half of the audiovisual recordings (30 clips). The 60 clips were recorded in a quiet room, using high quality equipment. The microphone was about 50 centimeters from the talker. Some of the sentences used in the recording were selected from the list in Munro and Derwing (1995) and others were created by the researcher. The list of sentences can be seen in the Appendix 1 section.

The 60 clips were divided into two conditions, 30 videos where the speaker wears the surgical MASK, and 30 videos with NO MASK. In each video the speaker showed one of the three emotions: happy, neutral, or angry. These short clips lasted from 2 to 3 seconds, and the participants could watch them only once. To increase the difficulty of the linguistic task, a multi-talker babble noise was added to the input with a resolution of +3 dB. The volume of the multi-talker babble was 3 decibels quieter than the utterances spoken by the talker, which was not uncomfortably difficult for listeners. However, only one participant commented that 2 videos were inaudible to her. Two participants reported difficulties with playing the video, their answers were taken into account.

Each trial consisted of a video presentation, an emotion rating, and a sentence transcription task. The participants could watch each video only once. After they watched a video, participants had three buttons (happy, neutral, angry) to choose from to identify the emotion portrayed in the video. Then, in the same screen, participants typed in a box what they understood as accurately as they could for each sentence. The 60 trials were randomly presented by the program (Qualtrics). Every participant watched 60 sentences, performing the same task for every video they watched. After the 60 clips, participants were asked to complete an anonymous short language questionnaire seen in the Appendix 2 section. They were asked about their age, nationality, places they have lived in, and languages spoken by them and by their primary caregivers.

Due to some COVID restrictions, this experiment could not be performed in person. In a regular setting, we would have called the participants to attend the phonetics lab to record their participation. The participants would have sat at a desktop computer, wearing headphones, and the researchers would have requested to turn their cell phones off to avoid distractions. Instead, an email was sent to two specific groups of students at the University of Kentucky, the Latino Students Association (LSA) and the undergraduate students from the Linguistics department (summer courses) to perform this experiment virtually. The first group was not offered any incentive, such as extra credit or monetary compensation for their participation. The second group, the LIN students, was offered extra credit in one class after completing this study. After the participants read an email sent to their university accounts, they clicked on a link that took them to a Qualtrics page where they read a consent form, and they agreed on participating in this experiment. Then, participants were asked to check on equipment functionality such as headphones and volume and complete the main experiment. The time expected for this study completion was no more than 20 minutes.

2.1 Using the correct equipment for speech perception experiments

Some people who take surveys voluntarily might not care enough about the outcomes. The results also depend on the participants' predisposition and emotional state, the equipment they use, and the place they are. For this study, we have recommended the participants to use headphones. We have also advised them to use a laptop or desktop computer; however, participants could use any other personal device such as cell phone or tablet.

2.2 Experiment 1

Forty-nine participants from the LSA group voluntarily participated in this study, but only fully completed surveys were registered and only 42 surveys were considered. Participants did not report any previous history or diagnosis of speech, language, or hearing disorders. The participants were students from the University of Kentucky who volunteered to participate in this study. The students from the LSA were expected to have a background in both languages, Spanish and English; and they indeed documented to have had contact with both languages for years. Their ages varied from 18 and over. From this group, 10 respondents were male and 32 females. As for race/ethnicity, 19 students described themselves as Hispanic, 12 as Latin, 2 as White, 1 as African American, and 8 as other (Afro-Caribbean, Latina and Hispanic, White/Hispanic, African American/Hispanic, mixed, and N/A). These students with Spanish and English background were contacted thanks to Ruth Gonzalez, community specialist of the Latino Students Association (LSA), 2021 from the University of Kentucky. This group of students were invited to participate voluntarily with no other incentive than supporting the Latin/Hispanic community at the University of Kentucky.

According to the language questionnaire, all the participants reported contact with Spanish and English for years and by influence of their primary caregivers. The Participants from the LSA have been classified in three groups, based on Regalado et al., (2019), and according to their language background: a) L1 refers to monolingual (English) speakers, 23 participants were counted. b) L2, speakers of English as their second language or late bilinguals, 12 participants were counted. c) L1-2 Speakers who grew up speaking both languages (English Spanish) and report the same number of years of its use; in this group, 7 people were counted.

2.3 Experiment 2

The experiment was directed to L1 English speakers, students from the department of Linguistics; this group has been referred as LIN group. Their instructors offered them extra credit in one assignment to motivate them to participate in this study. However, only 10 participants volunteered from this pool, 8 females and 2 males. 9 students reported their ethnicity/race as White, and 1 as African American. Their ages ranged from 18 to 54 years old. The 10 participants indicated that English was their first language. Two participants indicated they have spoken Spanish for 5 and 6 years, and one reported she has spoken Chinese for 9 years. The same input and the same procedure were applied to this group. Participants received the invitation via email, they agreed to participate in this study and completed the transcription task, the emotions rating, and the language questionnaire.

One of the differences between the LIN group and the L1 group (from the LSA group) is the caregivers' languages background. The LIN students indicated that their caregivers were monolingual English speakers, while most of the L1 group reported that their primary caregivers spoke English only or Spanish and English. Therefore, the L1 group has had more contact with both languages due to monolingual or bilingual caregivers.

CHAPTER 3. RESULTS

3.1 Experiment 1

3.1.1 Transcription accuracy

To interpret the results, there exist different ways to analyze speech perception in noise (SPIN) tests and transcription accuracy tests. For example, in SPIN tests, Smith and Fogerty (2017) analyzed how native listeners tend to reconstruct missed words from perceived pieces of information. They studied how native speakers substituted words for similar semantically and syntactically possible words. Smith and Fogerty also stated that more errors and gaps occur in SPIN tests compared to tests with no noise. Another method of calculation utilizes binary codes to assign correct or incorrect scoring (Li & Loizou, 2008). Zinszer, et al., (2019) examined error rates in whole utterances, contents words, and morphosyntactic levels. Needleman and Wunsch (1970) proposed a dynamic alignment algorithm that works with "a match award, a mismatch penalty, and a gap penalty."

There exist different ways to code transcription in speech perception studies. For example, McGowan (2015) used a Python script and set a Boolean "IsCorrect" variable to true for the matching final word. To keep analysis simple and to allow for comparison to other speech perception results (e.g., McGowan, 2015) the transcription response in this study was coded on the basis of the final, target word only as "Correct" or "Incorrect." Additionally, some morphosyntactic errors have been elided, cases where the transcriber used a different inflectional syntax in the word or used a semantically nearly equivalent word (e.g. "addicting" for *addictive*). Moreover, typographical errors (e.g. "briks" for *bricks*) have also been accepted and coded as "Correct." On the other hand, those participants who reported "*n/a," "I didn't understand"* and *it was inaudible* were coded as "Incorrect." Finally, the following replies *"the video glitched"* or "*the video did not play"* were separated, and the mean, mode, and median of different conditions were coded accordingly the new total number obtained without taking those responses into account.

The LSA group consisted of 42 participants who were classified in 3 groups, in the L1 group there were 23 participants, in the L1-2 group there were 7 participants, and in the L2 group there were 12 participants. The following tables and figures have been calculated based on the three different groups for more specific information. Some of the figures have also been created based on the total count of the participants' responses in the MASK and the total of responses in the NO MASK condition.

The following Table 3.1.1 displays the arithmetic mean of the congruence in the transcription task across the three different groups, the L1, L1-2, and L2 in the MASK and NO MASK condition. The mean of the congruent MASK condition by L1s corresponds to 27 followed by the mean error of 2.83. The mean of the congruent NO MASK condition of this group is 28.5 with a mean error of 1.5. The L1-2 group, in the MASK condition, has a mean of transcription accuracy of 27.9 and the mean error of 2.1, while in the NO MASK condition, it has an accuracy mean of 28.7 and mean error of 1.3. The L2 group has an accuracy mean of 23 followed by the mean error of 7, in the MASK condition. In the NO MASK condition, the L2 group has a mean of transcription accuracy of 26.8 and the mean error of 3.2. The results show less errors in the NO MASK condition across the three groups.

Table 3.1.1

Mean of Congruent Transcriptions and Errors by the LSA Group

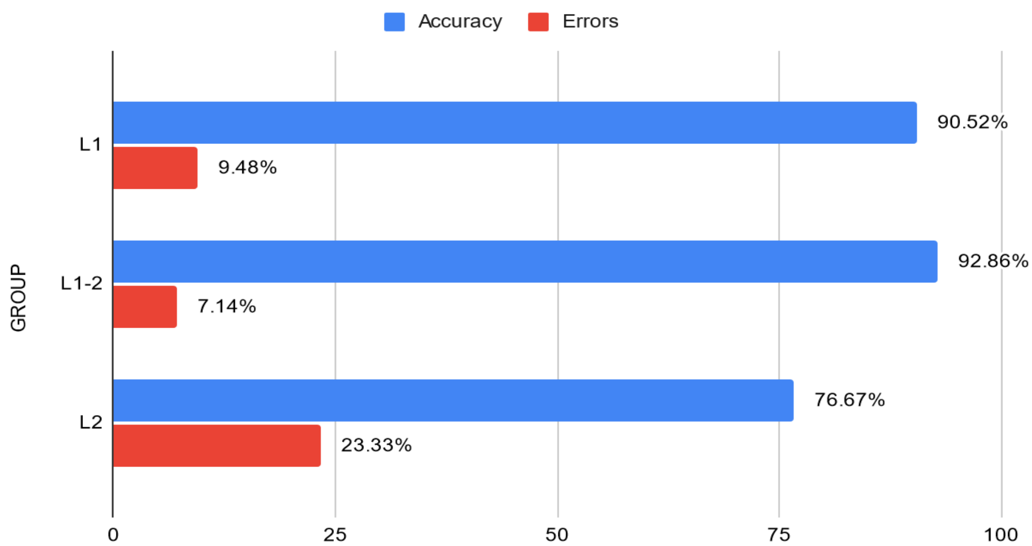
Group	Congruent MASK	Error MASK	Congruent NO MASK	Error NO MASK
L1	27	2.83	28.5	1.5
L1-2	27.9	2.1	28.7	1.3
L2	23	7	26.8	3.2

Note. Stimulus MASK and NO MASK for the three subgroups.

Figure 3.1.1 shows the percentages of accuracy in the transcription task by the three different groups, L1, L1-2, and L2s in the MASK condition. All 42 participants listened to and transcribed 30 sentences. In the L1 group, 23 participants have transcribed the 30 sentences in the MASK condition; however, 4 sentences were eluded from the total count due to the answer of two participants, "the video didn't play" and "the video glitched." It has been obtained a total of 686 transcriptions from this group. The L1 group obtained 90.52% accuracy corresponding to 621 accurate responses, 9.48% of errors which corresponds to 65 errors. The accuracy of L1-2 group, with 7 participants, reached 92.86% which represents 195 accurate answers and, 7.14% of errors which represent 15 errors. In the L2 group, with 12 participants, the total of sentences was 360. This group obtained 76.67% accuracy which represents 276 accurate transcriptions, and 23.33% of errors that corresponds to 84 errors. The percentage of accuracy of the L1-2 group shows a higher accurate perception rate compared to the two other groups. The L2 group obtained the lowest scores in accuracy in the transcription task MASK condition.

Figure 3.1.1

Percentages of the Transcription Task by the LSA Group. MASK Condition

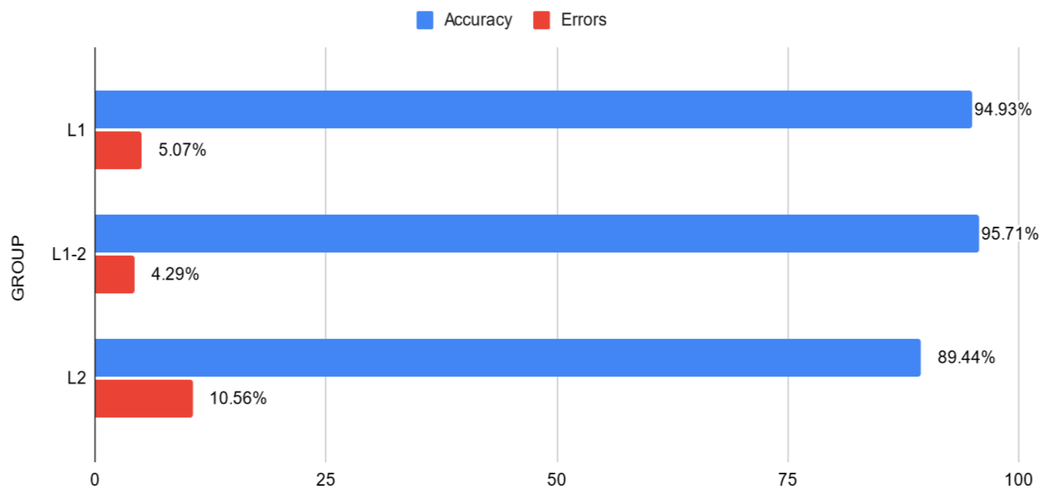


Note. Transcription task specified for the three subgroups of the LSA group.

The following Figure 3.1.2 shows the percentages of accuracy in the transcription task across the three groups, L1, L1-2, L2 in the NO MASK condition. The L1 group obtained 94.93% accuracy which represents 655 accurate transcriptions. This group obtained 5.07% of errors corresponding to 35 errors. The L1-2 group reached 95.71% accuracy that corresponds to 201 accurate responses and 4.29% of errors which represents 9 errors. Finally, the L2 group was 89.44% accurate that represents 322 accurate responses with 10.56% of errors which corresponds to 38 errors. As shown in Figure 3.1.1 and in Figure 3.1.2, the L1-2 group was more accurate in the transcription task compared to the L1 and L2 groups. Additionally, a chi-square test of independence was performed to examine the relation between the speech transcription with MASK and NO MASK and L1, L2, L1-2 speakers. However, the relation between these variables was not significant, X^2 (2, N = 42) = 1.279, p = .527553

Figure 3.1.2

Percentages of the Transcription Task by the LSA Group. NO MASK Condition



Note. Transcription task specified for the three subgroups of the LSA group.

Table 3.1.2 displays the MASK and NO MASK conditions with the mean, median, and mode of the LSA group followed by the mean error. In total, the LSA participants, in the MASK condition, obtained 26 accuracy transcriptions as the mean and 3.9 as the mean error. In the NO MASK condition, the mean was 28 with a mean error of 2. In the MASK condition, the median was 27 with 3 errors. In the NO MASK condition, the median was 28.5 with 1.5 errors. The mode in the MASK condition is 25 with 5 errors while the mode in the NO MASK condition is 29 with 1 error. The LSA group has obtained more accurate transcription results in the NO MASK condition than in the MASK condition.

Table 3.1.2

Total of Transcription Accuracy and Errors by the LSA Group.

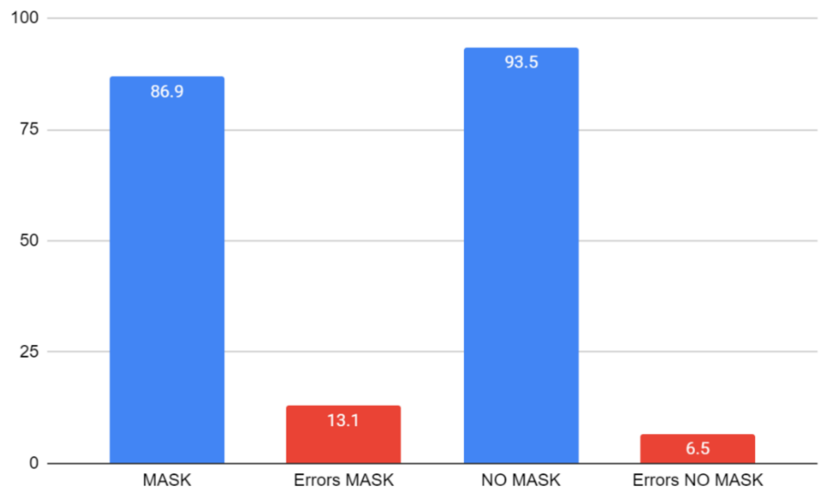
Measure	Congruent MASK	Error MASK	Congruent NO MASK	Error NO MASK
Mean	26	3.9	28	2
Median	27	3	28.5	1.5
Mode	25	5	29	1

Note. Measures of the stimulus MASK and NO MASK

The following Figure 3.1.3 shows that the LSA group obtained less accurate responses in the MASK condition with 86.9% which represents 1092 accurate responses; 13.1% of errors corresponds to 164 errors. Then, in the NO MASK condition participants obtained 1178 accurate transcriptions reaching 93.5% higher than the MASK condition. The 82 errors in the NO MASK condition reached 6.5%. This figure sums up how the NO MASK condition obtained more accurate results in the transcription task than the MASK condition, by the LSA group.

Figure 3.1.3

Percentages of Accuracy in the Transcription Task by the LSA Group



Note. Comparison of accuracy in the two stimulus, MASK and NO MASK.

3.1.2 Emotions accuracy

The tables and figures below show the analysis of the emotional prosody task according to the three groups, L1, L1-2, and L2 and divided in two parts, the emotion with MASK and the emotion with NO MASK. It has also been analyzed according to the three emotions, happy, neutral, and angry. Table 3.1.3 shows the mean of accurate responses across the three groups of the LSA participants in the MASK and NO MASK condition. The mean of the emotions' accuracy in the MASK condition by L1s corresponds to 26.13 followed by the error mean of 3.87. The mean in the NO MASK condition by L1s corresponds to 27.43 followed by the error mean of 2.57. The L1-2 group in the MASK condition has a mean of 26.43 and error mean of 3.57. The mean of this group in the NO MASK condition is 27.86 and error mean of 2.14. The L2 group has a mean of 25.58 and mean of error of 4.42. In the NO MASK condition, the mean of the L2 group is 26.5 and error mean of 3.5. The L2 group shows a slightly lower mean in the accuracy of emotions perception compared to the other two groups.

Table 3.1.3

Mean of Accuracy in the Perception of Emotions by the LSA Group

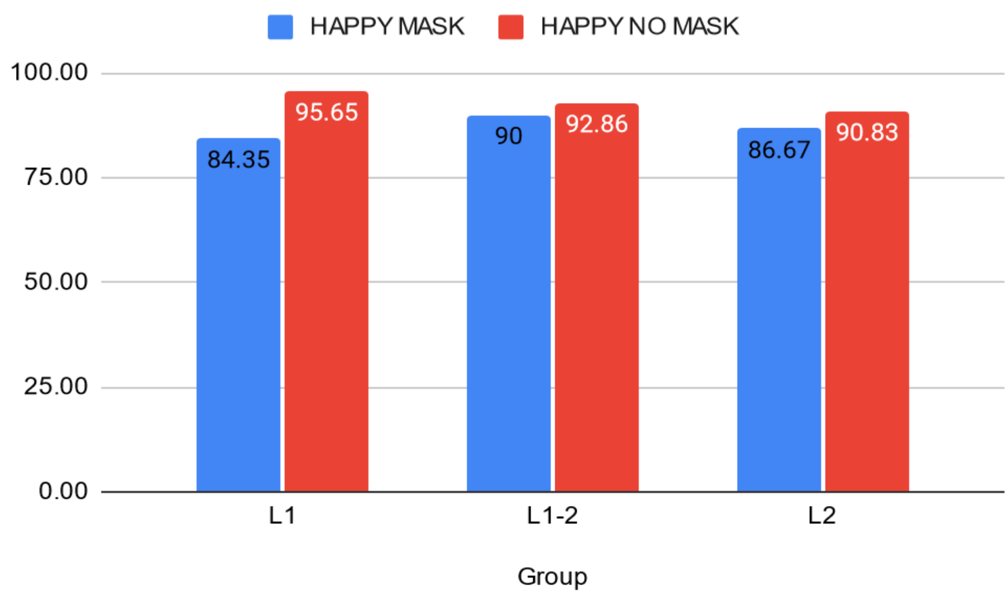
Group	Accuracy MASK	Error MASK	Accuracy NO MASK	Error NO MASK
L1	26.13	3.87	27.43	2.57
L1-2	26.43	3.57	27.86	2.14
L2	25.58	4.42	26.5	3.5

Note. Stimulus MASK and NO MASK for the three subgroups

Each emotion was also analyzed according to the accuracy obtained by the three groups, L1, L1-2, and L2 in the MASK and NO MASK condition seen in Figure 3.1.4. The HAPPY MASK emotion obtained 194 accurate responses that represents 84.35% of L1 participants. For the HAPPY NO MASK prosody, 220 responses were accurate which represents 95.65%. L1-2s obtained 90% that is 63 accurate responses in the HAPPY MASK condition and 92.86% that is 65 accurate responses in the HAPPY NO MASK condition. L2s reached 86.67% that is 104 accurate responses in the MASK condition, and 90.83% that is 109 accurate responses in the NO MASK condition. The percentages of the NO MASK condition are consistently higher than the MASK condition across all three groups. In the NO MASK condition, the percentages of L1 are higher than L1-2, which is also higher than the L2 group. However, the L1 group, in the HAPPY MASK condition, presents a lower number compared to the L1-2 and L2 groups.

Figure 3.1.4

Percentages of Accuracy in the HAPPY Prosody by the LSA Group



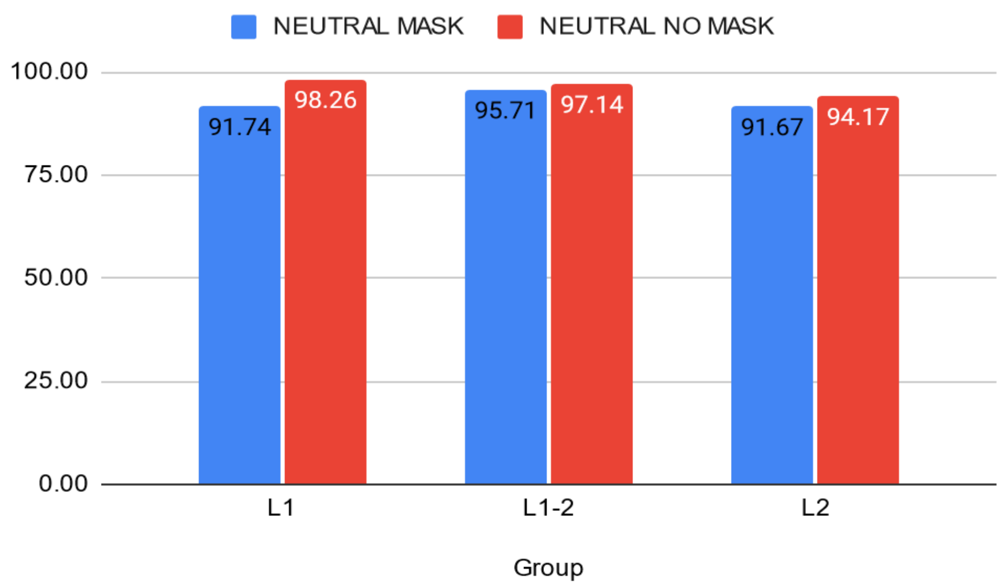
Note. Percentages of the three subgroups for the MASK and NO MASK stimuli

Figure 3.1.5 shows the NEUTRAL prosody percentages. The L1 participants obtained 211 accurate responses that corresponds to 91.74% accuracy in the perception of the NEUTRAL MASK stimulus. In the NEUTRAL NO MASK stimulus, participants obtained 226 accurate responses that is 98.26%. The L1-2 group obtained 67 accurate responses that is 95.71% accuracy in the MASK condition and 68 accurate responses that is 97.14% in the NO MASK condition. The L2 participants obtained 110 accurate responses that is 91.67% accuracy in the perception of the MASK stimulus and 113 accurate response that is 94.17% in the NO MASK stimulus. A similar pattern is seen in this figure, as it was seen in the HAPPY emotion; the NEUTRAL NO MASK condition obtained higher percentages than the NEUTRAL MASK condition across the three groups.

For the NO MASK condition, L1 is higher than L1-2, which is also higher than L2. However, for the MASK condition, the L1 group has a lower percentage than the L1-2 group; this is similar to the HAPPY emotion seen above. And the L2 group reached the lowest percentage across the three groups.

Figure 3.1.5

Percentages of Accuracy in the NEUTRAL Prosody by the LSA Group

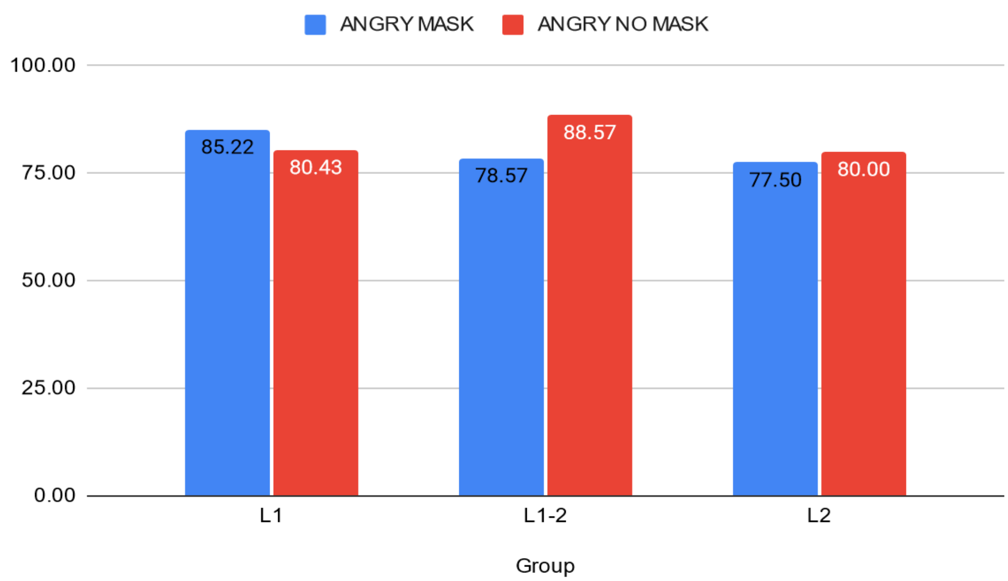


Note. Percentages of the three subgroups for the MASK and NO MASK stimuli

The accuracy of the ANGRY prosody is seen in Figure 3.1.6. L1 participants obtained 196 accurate responses that is 85.22% accuracy in the ANGRY MASK condition and 185 accurate responses that is 80.43% in the NO MASK condition. L1-2s obtained 55 accurate responses that is 78.57% accuracy in the MASK condition and 62 accurate responses that is 88.57% accuracy in the NO MASK condition. L2s obtained 93 accurate responses that is 77.50% accuracy in the MASK condition; and 96 accurate responses that is 80% accuracy in the NO MASK condition. The NO MASK condition was higher for the L1-2 and L2 groups compared to the MASK condition. However, the L1 group obtained a different percentage compared to the previous figures; the ANGRY MASK condition reached a higher percentage than the NO MASK condition.

Figure 3.1.6

Percentages of Accuracy in the ANGRY Prosody by the LSA Group



Note. Percentages of the three subgroups for the MASK and NO MASK stimuli

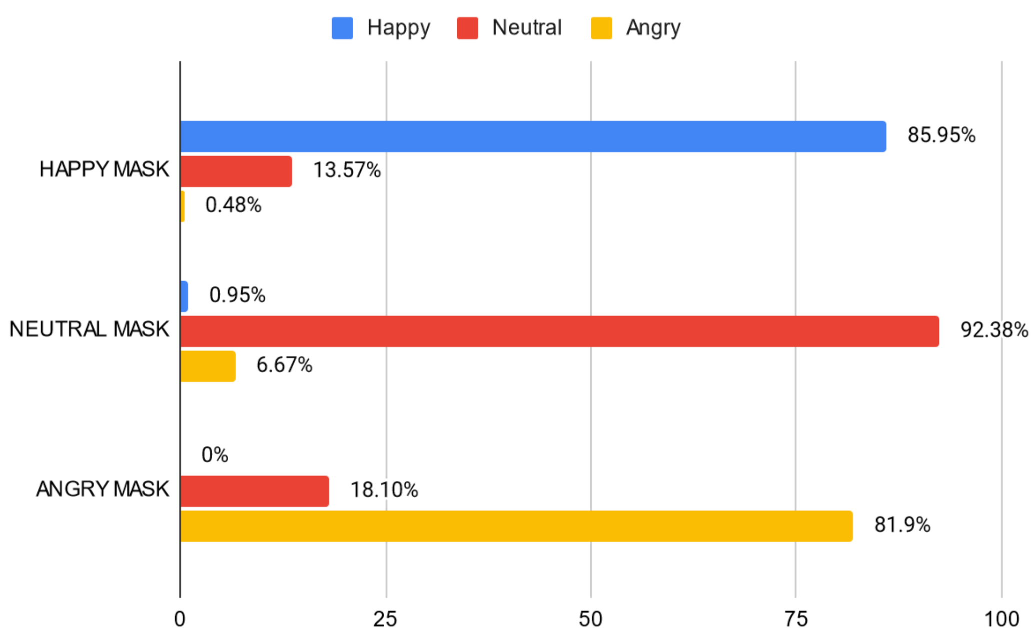
The following figures group all the LSA participants together and show the percentage of accuracy and errors they all obtained in the three different emotions. The numbers and percentages given came from the 42 participants who perceived and responded to the three groups of phrases with the different emotions, ten in each group (10 phrases HAPPY MASK, 10 phrases HAPPY NO MASK, 10 phrases NEUTRAL MASK, 10 phrases NEUTRAL NO MASK, 10 phrases ANGRY MASK, 10 phrases ANGRY NO

MASK). The overall number of responses in every group of emotions is 420 which corresponds to 100%.

Figure 3.1.7 shows the MASK condition where 361 responses were accurate in the perception of the HAPPY MASK prosody, this represents 85,95% of accuracy. This HAPPY MASK emotion was perceived as neutral, and it obtained 57 responses that represents 13.57%. This emotion also obtained 2 responses perceived as angry that is 0.48%. The perception of the NEUTRAL MASK prosody obtained 92.38% accuracy that corresponds to 388 accurate responses. The NEUTRAL MASK emotion was also perceived as happy with 4 responses that is 0.95%; and it was also perceived as angry with 28 responses that represents 6.67%. Finally, the perception of the ANGRY MASK prosody obtained 81,9% accuracy that corresponds to 344 accurate responses. This emotion was perceived as neutral, with 76 responses which represents 18.10%; and nobody perceived it as happy. The participants were more accurate in the perception of the NEUTRAL prosody followed by the HAPPY emotion, and then by the ANGRY emotion.

Figure 3.1.7

Percentages of the Accuracy in the Perception of Emotions. MASK Condition



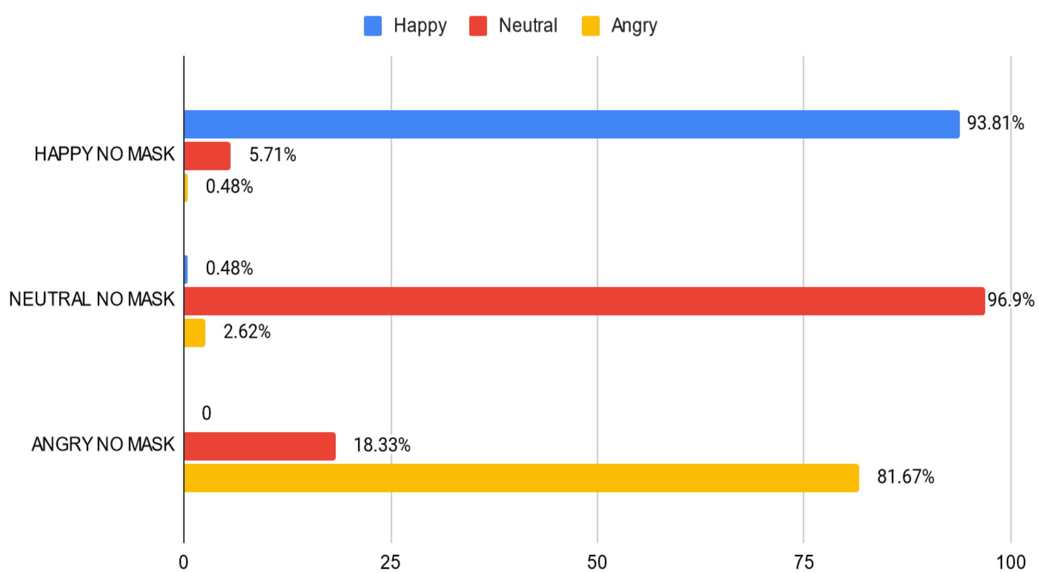
Note. Perception of the three emotions by the LSA group.

The next Figure 3.1.8 shows the three emotions in the NO MASK condition. The HAPPY NO MASK prosody obtained 394 accurate results that is 93.81%; this emotion was perceived as neutral 24 times that represents 5.71%, and 2 times as angry that represents 0.48%. The NEUTRAL NO MASK emotion was accurately perceived 407 times corresponding to 96.9%, there were 2 responses of participants who perceived the neutral emotion as happy, and this represents 0.48%; the neutral emotion was also perceived as angry two times that is 0.48%. The ANGRY NO MASK emotion was accurately perceived 343 times that is 81.67%, the angry emotion was also perceived 77 times as neutral which represent 18.33%; nobody perceived it as happy. The outcomes of the NEUTRAL NO

MASK prosody were higher than the HAPPY NO MASK which is also higher than the ANGRY NO MASK prosody. These outcomes are similar to the previous graphic where the three emotions obtained the same order in accurate perception, first NEUTRAL, second HAPPY, and third ANGRY. Interestingly, the results show that the ANGRY MASK stimulus was slightly better perceived than the ANGRY NO MASK stimulus.

Figure 3.1.8

Percentages of the Accuracy in the Perception of the Emotions. NO MASK Condition



Note. Perception of the three emotions by the LSA group.

According to these results, the emotion that was most accurately perceived was NEUTRAL NO MASK. Also, the HAPPY NO MASK stimulus was well perceived by participants. The ANGRY NO MASK stimulus was the least accurate perceived emotion by the LSA group. The percentage of the ANGRY MASK emotion is higher than the ANGRY NO MASK stimulus. In general for the two conditions, participants were more accurate in identifying the NEUTRAL prosody, followed by the HAPPY prosody, and then the ANGRY prosody. Additionally, the chi-square test of independence was performed to examine the relation between the perception of the three emotions and the MASK and NO MASK condition. However, the relation between these variables was not significant, X^2 (2, N = 42) = 0.7356, p = .692258.

Table 3.1.4 shows the mean, median, and mode of accuracy in perception of emotions with the conditions MASK and NO MASK. Overall, the perception of the emotions in the MASK condition had a mean of 26.02 with a mean of error of 3.98. While the perception of the emotions in the NO MASK condition had a mean of 27.24 with a mean of error of 2.76. In the MASK condition, the median was 28 with 2 errors; in the NO MASK condition the median was 29 with 1 error. Finally, the mode in the MASK condition was 30 and 0 errors, and the mode in the NO MASK condition was also 30 and 0 errors. The NO MASK condition displays more accurate results than the MASK condition.

Table 3.1.4

Total of Accuracy and Errors in the Emotions Rating by the LSA Group

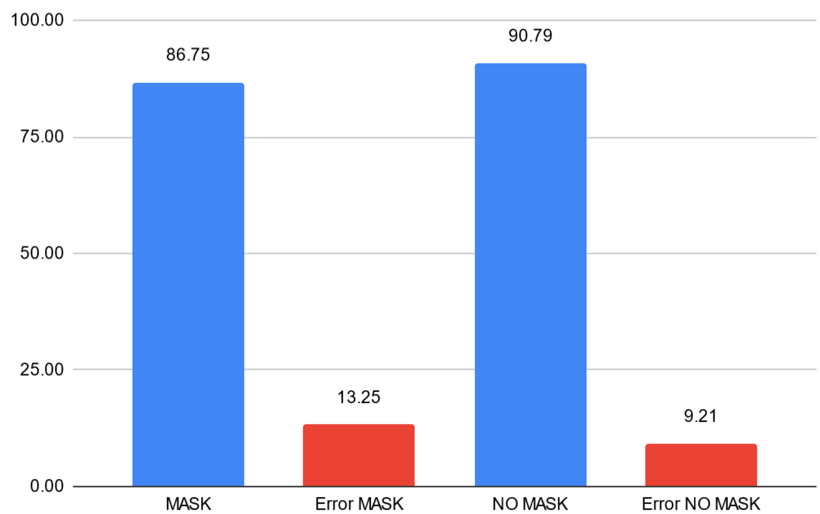
Measures	Accuracy MASK	Error MASK	Accuracy NO MASK	Error NO MASK
Mean	26.02	3.98	27.24	2.76
Median	28	2	29	1
Mode	30	0	30	0

Note. Measures of the MASK and NO MASK stimulus.

Finally, Figure 3.1.9 shows the percentages of accuracy in the perception of emotions in the two conditions MASK and NO MASK by the LSA group. The 42 participants listened and perceived to 30 emotions with MASK and 30 emotions with NO MASK. The MASK stimulus obtained 1093 accurate responses that represents 86.75% with 167 errors that is 13.25%. The NO MASK stimulus obtained in total 1144 accurate responses that is 90.79% and 116 errors which represents 9.21%.

Figure 3.1.9

Percentages of Accuracy in the Perception of Emotions. LSA Group



Note. Percentages in the MASK and NO MASK stimulus

3.2 Experiment 2

The results from the LIN group (10 students from the department of Linguistics) differ in number with participants with experiment 1. This group of students were offered extra credit in one of the courses; however, there was a low involvement of students. The LIN participants transcribed 30 sentences with the MASK condition, and 30 with the NO MASK condition; in total, 300 transcriptions were obtained in each condition. The way these results were coded shares the same procedure as experiment 1. For the transcription task, the match of the last content word of the utterance was coded as "Correct" or "Incorrect." Additionally, typographical errors (e.g., "oxygn" for *oxygen*) have also been accepted and coded as "Correct." On the other hand, some answers reported as unintelligible or comments such as *I couldn't hear* were coded as "incorrect". The outcomes of Experiment 2 are shown in the tables and figures below.

3.2.1 Transcription accuracy

The results obtained from the LIN group is seen in Table 3.2.1 below where the stimulus with MASK obtained 26.9 congruent transcriptions on average accompanied by the mean error of 3.1. Their median was 27 in congruent transcriptions followed by the median of errors of 3. The mode had a transcription accuracy of 27 and the mode of 3 errors. The NO MASK stimulus obtained an average of 27.8 congruent transcriptions accompanied by the mean of errors of 2.2. Their median was 28.5 in congruent transcriptions followed by 1.5 errors, and the mode shows accuracy of 30 congruent transcriptions and 0 errors. It is seen that the NO MASK condition obtained more accurate responses compared to the MASK condition.

Table 3.2.1

Total of Accuracy in the Transcription Task and Errors by the LIN Group

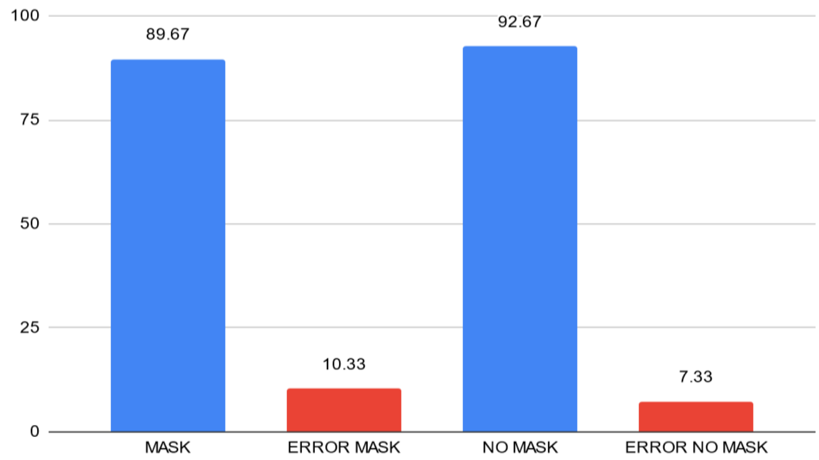
Measure	Congruent MASK	Error MASK	Congruent NO MASK	Error NO MASK
Mean	26.9	3.1	27.8	2.2
Median	27	3	28.5	1.5
Mode	27	3	30	0

Note. Measures for the MASK and NO MASK stimulus.

The following Figure 3.2.1 shows the percentages of accuracy in the transcription task in the two conditions, MASK and NO MASK. Participants obtained 269 accurate transcription that represents 89.67% in the MASK condition with 31 errors that is 10.33%. The accuracy in the NO MASK condition obtained 278 accurate responses that represents 92.67% with 22 errors that is 7.33%. This figure shows that the NO MASK condition obtained higher transcription accuracy also with monolingual speakers.

Figure 3.2.1

Percentages of Accuracy in the Transcription Task by the LIN Group



Note. Percentages for the MASK and NO MASK stimulus.

According to this data, the accuracy of transcriptions with the NO MASK stimulus has an advantage over the MASK stimulus. Similar to the first experiment, the NO MASK condition shows better intelligibility of speech.

3.2.2 Emotions accuracy

The analysis of the emotional prosody was coded according to the three emotions and divided in two parts, condition with MASK and NO MASK. Table 3.2.2 displays the mean of accuracy in the emotional prosody rating where the MASK condition presents the accuracy mean of 20 and 10 errors. The NO MASK condition presents a mean of 22 with a mean of error of 8. The median in the MASK condition is 21.5 with 8.5 errors. In the NO MASK condition, the median is 24 with 6 errors. The mode in the MASK condition is 10 with 20 errors while the mode in the NO MASK condition is 28 with 2 errors. The NO MASK condition once again shows an advantage over the MASK condition.

Table 3.2.2

Total Accuracy in the Emotional Prosody Task by the LIN Group

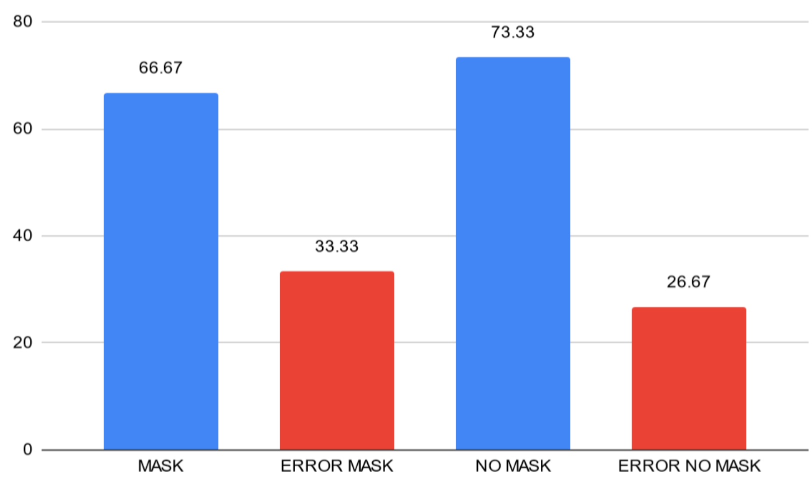
Measure	Accuracy MASK	Errors MASK	Accuracy NO MASK	Errors NO MASK
Mean	20	10	22	8
Median	21.5	8.5	24	6
Mode	10	20	28	2

Note. Measures of the MASK and NO MASK stimulus.

Figure 3.2.2 shows the percentages of accuracy in the perception of emotions with the MASK and NO MASK condition and their errors. In the MASK condition, participants obtained 200 accurate responses that represents 66.67% with 100 errors that is 33.33%. The NO MASK condition obtained 220 accurate responses which represents 73.33% and 80 errors that is 26.67%. The NO MASK condition obtained more accurate results than the MASK condition.

Figure 3.2.2

Percentages of Accuracy in the Perception of Emotions. LIN Group

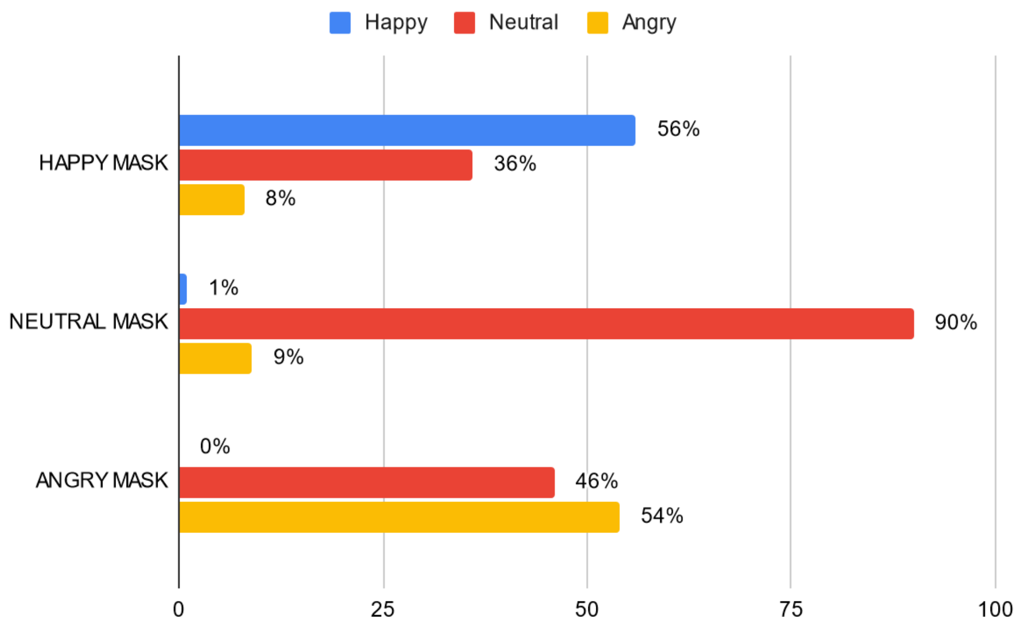


Note. Percentages for the MASK and NO MASK stimulus.

Figure 3.2.3 shows the percentages of accuracy of the emotional prosody in the MASK condition. First, the HAPPY MASK prosody obtained 56 accurate responses that represents 56%, this emotion was also perceived as neutral 36 times that is 36%, and 8 times as angry that is 8%. The perception of the NEUTRAL MASK prosody obtained 90% accuracy, 9% was perceived as angry, and 1% was perceived as happy. The ANGRY MASK prosody obtained 54% of accuracy, 46 times were perceived as neutral that is 46%, and 0% as happy. It is seen again that the NEUTRAL prosody was much better perceived by participants than the HAPPY and ANGRY emotion. In the three individual prosodies, the NEUTRAL emotion was significantly present, participants constantly opted for choosing this emotion.

Figure 3.2.3

Percentages of Accuracy in the Perception of Emotions with MASK.

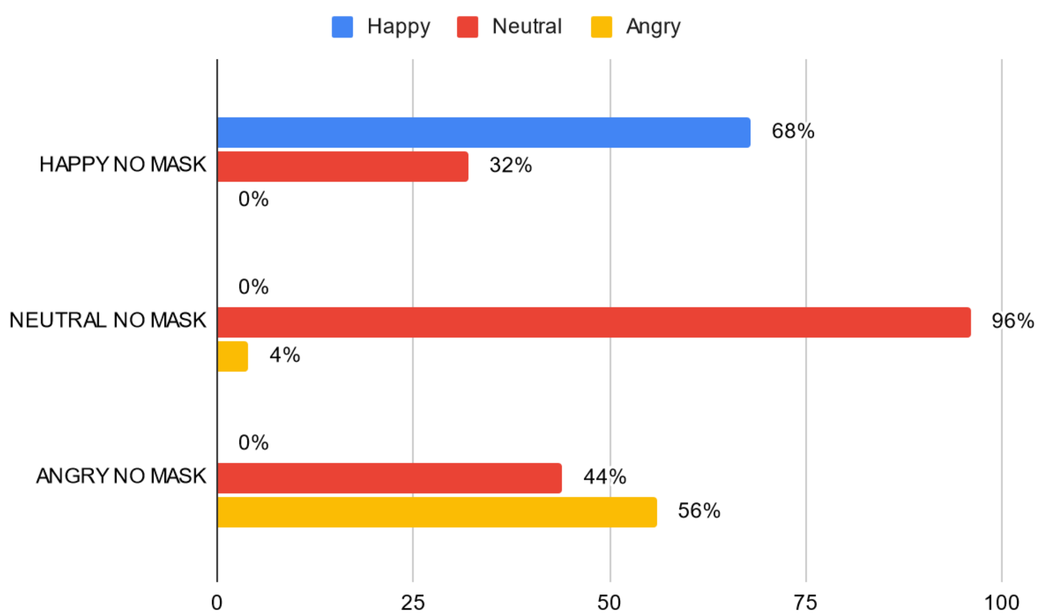


Note. Perception of the three emotions by the LIN group.

Figure 3.2.4 shows the percentages of the emotional prosody in the NO MASK condition. Participants accurately perceived the HAPPY NO MASK prosody in 68%, 32% was perceived as neutral, and 0% as angry. The NEUTRAL NO MASK emotion was accurately perceived in 96%, 4% of the group perceived it as angry, and 0% as happy. For the ANGRY NO MASK stimulus, 56% of participants perceived it accurately, 44% perceived it as neutral, and 0% as happy. The NO MASK condition presents more accurate responses by the LIN group. Once again, the NEUTRAL prosody is significantly present and strongly perceived by the participants in the three intended emotions.

Figure 3.2.4

Percentages of Accuracy in the Perception of Emotions with NO MASK



Note. Perception of the three emotions by the LIN group.

On average, the three emotions were better perceived in the NO MASK stimulus as seen in Figure 3.2.2. Similar to experiment 1, participants were more accurate in identifying the NEUTRAL prosody, followed by the HAPPY prosody, and then by the ANGRY prosody. Another interesting point is the high rate of perception of the NEUTRAL emotion instead of the intended angry stimulus. The NEUTRAL prosody has been chosen by participants repeatedly across the two conditions, MASK and NO MASK, and across the three emotions. Additionally, the chi-square test of independence was performed to examine the relation between the three emotions and the MASK and NO MASK condition in the LIN group. However, the relation between these variables was not significant, X^2 (2, N = 10) = 0.4398, p = .802592.

CHAPTER 4. DISCUSSION

The extra barriers implemented in this experiment such as the surgical mask and the multi-talker babble noise might have contributed to obtaining higher number of errors in the NO MASK stimulus in the transcription task and emotional rating of experiment 1 and experiment 2. It is important to remember that participants' replies as *the video didn't play*, or *it glitched* were avoided in the coding process, and new totals were used to obtain the mean and percentages. The use of the surgical mask enlarged the number of errors in both tasks, lexical intelligibility and emotions rating. Surgical masks seem to attenuate some sounds and minimize the visual information that affects the perception of different emotional cues and the comprehension of the linguistic message. However, the chi square of independence showed that there was no significant association between the MASK / NO MASK variable and the emotional prosody in either group, LSA and LIN.

The language background can be a factor that affects speech perception. Those L2 speakers who acquired the language later in life present lower accuracy in the transcription task in the MASK and NO MASK condition compared to the other two groups, L1 and L1-2 speakers. These groups have been formed to show more granulated information about how different language backgrounds may affect speech perception. However, these subgroups created to classify the language backgrounds of participants have not been a meaningful number due to a low involvement of LSA students. Likewise, the chi-square of independence showed that there was no significant association between the MASK / NO MASK variable and the classification of language background (L1, L1-2, L2).

The low involvement of students and some other variables that could have altered the results of this experiment can be the use of not adequate equipment, the place where the experiment was developed, the predisposition and mental state of the participants at the moment of taking the online experiment. Additionally, working with an anonymous online project does not give a 100% accurate information. Some participants may not answer all the questions or avoid very important information that could have been inquired if the experiment were applied in person.

CHAPTER 5. CONCLUSIONS

The perception of the emotional prosody mode and the congruence of the transcription task highlights a higher accuracy in the NO MASK stimulus as predicted. However, the number of participants involved in this study and their results in the transcription task do not give significant outcomes that could highlight the variables as affecting variables. Future studies with more participants could support Llamas et al.'s (2008) findings about the loss of speech intelligibility due to lack of visual information. It is also important to do more research and add evidence for reduction of accuracy of the emotional prosody perception by the surgical mask in different groups with different language backgrounds. Interestingly, the L2 speakers obtained the lowest percentages and mean in the perception of emotions in both conditions MASK and NO MASK.

Further, results from this work show that some people associated the "angry" emotion with "neutral," these results differ from Kim and Sumner's (2017) findings, with auditory-visual priming, where some people associated the "neutral" prosody with "happy." Communication can be altered in the presence of a surgical mask, and the comprehension of the linguistic message including emotional prosody can change to slightly different degrees.

A second language background can be another factor that influences the way the information is received. A person's degree of exposure to the language of the utterance, and age of this L2 acquisition are some variables that intervene with increasing or decreasing the performance in communication (Tabri et al., 2011). In this work, the L2 speakers performed lower in both tasks, transcription and emotional rating compared to the two other groups. However, the outcomes presented in this study were not significant. The classification of participants in different groups can help identify the need some people have when communicating using a language that is not their first language. Knowing about the difficulties different groups can have with languages, can help us recognize the need of using a clearer and slower pronunciation when communicating.

The LIN students performed better in the NO MASK condition. The mask used for the transcription and emotion tasks affected the outcomes of both groups, the LSA and the LIN group. Nevertheless, the statistical test of independence indicated that the outcomes were not significant. In addition to these results, this study also showed that speech perception experiments via online can present more difficulties when classifying and understanding the data provided by participants because of the lack of interaction between investigators and participants. The researchers need to use their intuition more and trust in participants' answers, specifically in responses such as “I didn’t understand” or “the video did not play” which might have been just the result of external distractions. Performing in-person studies with the correct equipment and in the right place, for example a booth, might help participants avoid external distractions and increase their accurate outcomes.

APPENDICES

APPENDIX 1. SENTENCES FOR THE INPUT

1. Elephants are big animals. (NEUTRAL MASK)
2. March has thirty-one days. (NEUTRAL MASK)
3. Exercise is good for your health. (NEUTRAL MASK)
4. Japan is a wealthy country. (NEUTRAL MASK)
5. Some teenagers like rock and roll. (NEUTRAL NO MASK)
6. Some people love to eat chocolate. (HAPPY NO MASK)
7. Some people keep dogs as pets. (NEUTRAL NO MASK)
8. Young children can be very noisy. (NEURTRAL MASK)
9. Some roses have a beautiful smell. (ANGRY NO MASK)
10. Cats like to chase mice. (ANGRY MASK)
11. Red and green are colours. (NEUTRAL NO MASK)
12. Many houses are made of bricks (ANGRY MASK)
13. Italy is a country in Europe. (HAPPY MASK)
14. Many people drink coffee. (HAPPY NO MASK)
15. The American flag has stars. (NEUTRAL MASK)
16. Gold is a valuable metal. (NEUTRAL MASK)
17. You can buy a burger at McDonalds. (ANGRY MASK)
18. Ships travel on the water. (NEUTRAL NO MASK)
19. A bear is a kind of mammal. (ANGRY NO MASK)
20. The sun comes out in the day. (NEUTRAL NO MASK)
21. The inside of an egg is yellow. (HAPPY MASK)

22. Apples grow on trees. (ANGRY NO MASK)

23. The president lives in Washington. (NEUTRAL NO MASK)

24. Some people cook at home. (HAPPY NO MASK)

25. Most houses have electricity. (HAPPY MASK)

26. Some bats keep pests away. (ANGRY NO MASK)

27. Caffeine is a natural pesticide. (ANGRY MASK)

28. Some plants are edible. (HAPPY NO MASK)

29. Bamboo is the fastest growing plant. (ANGRY NO MASK)

30. Oceans produce earth's oxygen. (HAPPY NO MASK)

31. You need a coat for winter. (ANGRY MASK)

32. Sharks are large predators. (NEUTRAL NO MASK)

33. Some birds have good memory. (ANGRY NO MASK)

34. Some people are afraid of technology. (HAPPY NO MASK)

35. That beet looks fresh. (ANGRY NO MASK)

36. You can recycle some plastics. (NEUTRAL NO MASK)

37. Ducks can swim. (ANGRY MASK)

38. Chocolate can make you happy. (ANGRY MASK)

39. You can eat some flowers. (ANGRY MASK)

40. A year has twelve months. (HAPPY NO MASK)

41. Broccoli is actually a flower. (ANGRY NO MASK)

42. Teachers work in schools. (NEUTRAL NO MASK)

43. A break-up causes physical pain. (ANGRY NO MASK)

44. Love can be addictive. (HAPPY MASK)

45. E-mail is fast. (HAPPY MASK)

46. You find books in libraries. (ANGRY NO MASK)

47. Nurses work in hospitals. (HAPPY MASK)

48. Vaccines prevent diseases. (HAPPY MASK)

49. The earth travels around the sun. (ANGRY MASK)

50. There are seven continents. (ANGRY NO MASK)

51. Some flights are cheap. (HAPPY MASK)

52. A duck ate a pear. (ANGRY MASK)

53. There is fur on my seat. (HAPPY MASK)

54. A pain in your back is bad. (HAPPY NO MASK)

55. Some people like pets. (HAPPY NO MASK)

56. I see someone sing. (NEUTRAL NO MASK)

57. The president lives in Washington. (NEUTRAL NO MASK)

58. Gold is a valuable metal (HAPPY MASK)

59. Many houses are made of bricks (NEUTRAL MASK)

60. Young children can be very noisy. (NEUTRAL MASK)

APPENDIX 2. LANGUAGE QUESTIONNAIRE

Experiences with different languages can influence the way we talk and the way we perceive speech. Please, provide the following background information that will be helpful to researchers who study language.

1. Age: _____________

2. Gender

__ Male

__ Female

__ Other

3. Race or Ethnicity (check all that apply):

__ White

__ African American

__ Hispanic Latino

__ Asian

__ Hawaiian/Pacific Islander

__ Native American

__ Other _____________

4. Birthplace:

Country:___________________

5. Please indicate where, and how long you have lived in locations other than your birthplace:

Place 1. _____________________________

Place 2. _____________________________

Place 3. _____________________________

6. Which language(s)were spoken by your primary caregivers? Check all that apply:

	Caregiver1	Caregiver 2
English:	____	____
Spanish:	____	____
Other(s):	____	____

7. Which languages do you speak? Indicate for how many years:

English ______________________________

Spanish ______________________________

Other(s): ______________________________

REFERENCES

Baese-Berk, M.M., Bradlow, A. R., & Wright, B. A., (2013). Accent-Independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 133*(3), EL174-EL180.

Baese-Berk, M.M., McLaughtlin, D. J., & McGowan, K. B. (2020). Perception of non native speech. *Language and Linguistic Compass,* 2020 1-20.

Borneman, J. (2009). Veiling and women's intelligibility. *Cardozo Law Review, 30*(6) 2745 http://web.a.ebscohost.com.ezproxy.uky.edu/ehost/pdfviewer/pdfviewer?vid=1&sid=3397a430-33eb-4b3e-a6e2-c0f8ae4a0db1%40sdc-v-sessmgr02

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25*(5). 975-979.

Correy, R. M., Jones, U., & Singer, A. C. (2020). Acoustic effects of medical, cloth, and transparent face masks on speech signal. *Acoustical Society of America.* https://doi.org/10.1121/10.0002279

Docherty, G.J., & Foulkes, P. (2014). An evaluation of usage-based approaches to the modelling of sociolinguistic variability. *Lingua 142.* 42-56.

Ekman, P. & Friesen, W. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion*, 10(2). 159-168.

Enoch, J., McDonald, L., & Jones, L. (2019). Evaluating whether sight is the most valued sense. *Jama Ophthalmol, 137*(11), 1317-1320.

Fecher, N. (2014). *Effects of forensically-relevant facial concealment on acoustic and perceptual properties of consonants*. (Publication No. 1784059976) Doctoral dissertation, University of York. ProQuest Dissertations Publishing

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5).

Goldin, A., Weinstein, B., & Shiman N. (2020). How do medical masks degrade speech reception? *The Hearing Review.* Retrieved https://www.hearingreview.com/hearing-loss/health-wellness/how-do-medical-masks-degrade-speech-reception

Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory & Cognition, 12*(6), 568-574.

Johnson, J.S., & Newport, E. L. (1989). Critical period effect in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.

Johnstone, B., & Bean, J.M. (1997). Self-expression and linguistic variation. *Language in Society*, *26*(2), 221-246.

Jones, D. (2019, May). GP facing the sack for asking a Muslim woman to lift her veil so he could hear what was wrong with her little girl says he´s ´bowled over´ by public´s support after 59,000 sign petition. *The Daily Mail.com*. Retrieved March 10, 2021, from https://www.dailymail.co.uk/news/article-7069015/GP-asked-mother-surgery-lift-veil-no-longer-career.html

Kang, O. & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology, 28*(4), 441-456.

Kilman, L., Zekveld, A., Hallgren M., & Ronnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Frontiers in Psychology, 5,* 1-9.

King, S., & Sumner, M. (2014). Voices and variants: Effects of voice on the form-based processing of words with different phonological variants. Preceding of the Annual Meeting of the Cognitive Science Society, 36(36) California Digital Library

Kim, S. K., & Sumner, M. (2017). Beyond lexical meaning: The effect of emotional prosody on spoken word recognition. *The Journal of the Acoustical Society of America*, 142(1), EL49-EL55. https://doi-org.ezproxy.uky.edu/10.1121/1.4991328

Llamas, C., Harrison, P., Donnelly, D., & Watt, D. (2008). Effects of Different Types of Face Coverings on Speech Acoustic and intelligibility. *York Papers in Linguistics Series 2*. Issue 9. 80-104.

Leather, J. H. (1999). Second language speech research: An introduction. *Language Learning*. *49*(1). 1-56

Lenneberg, E. H. (1967). *Biological foundations of language*. New York, Wiley

Li, N. & Loizou, P. (2008). Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction. *The Journal of the Acoustical Society of America, 123*(3). 1673-1682.

Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of second-Language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research, 40*(3). 686-693.

McGowan, K. (2015). Social expectation improves speech perception in noise. *Language and Speech, 58*(4). 502-521.

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

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3). 289-306.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology, 48*(3). 443-453.

Nygaard, L. C., Sommers, M.S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*(1) 42-46.

Ochs, E. (1992). Indexing gender. In A. Duranti & C. Goodwin. (Eds.), *Rethinking context: Language as an interactive phenomenom* (pp. 335-358). Cambridge University Press.

Office of Students Financial Aid and Scholarships. (2021, February 20). *International students (apply).* https://www.uky.edu/financialaid/international.htm

Parikh, G., & Loizou, P. (2005). The influence of noise on vowel and consonant cues. *The Journal of the Acoustical Society of America, 118*(6), 3874-3888.

Potter, M.C., So, K., Eckardt, B. V., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior. 23*(1), 23-38. https://doi.org/10.1016/S0022-5371(84)90489-4

Regalado, D., Kong, J., Buss, E., & Calandruccio, L. (2019). Effect of language history on sentence recognition in noise or two-talker speech: Monolingual, early bilingual, and late bilingual speakers of English. *American Journal of Audiology, 28*(4), 935-946.

Saigusa, J. (2017). The effects of forensically relevant face coverings on the acoustic properties of fricatives. *Lifespans & Styles, 3*(2), 39-52.

Smith, K. G., & Fogerty, D. (2017). Speech recognition patterns for steady-state noise and interrupted speech. *The Acoustical Society of America, 142*(3). EL306-EL312.

Sueyoshi, A., & Hardison, D. (2005). The role of gestures and facial cues in second language comprehension. *Language Learning, 55*(4). 661-699.

Schwartz, J., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, *93*(2). B69-B78.

Tabri, D., 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.ezproxy.uky.edu/10.3109/13682822.2010.519372

Toscano, J.C., & Toscano, C. M. (2021). Effects of face masks on speech recognition in multi-talker babble noise. *Plos One*, *16*(2). e0246842 https://doi.org/10.1371/journal.pone.0246842

Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio-visual speech perception is special. *Cognition 96*(1). B13-B22.

Van Edwards, V. (2021, January 20). *How to read faces… Even when everyone is wearing a mask*. Scienceofpeople. https://www.scienceofpeople.com/masks/

Van Engen, K. J. (2010). Similarity and familiarity: Second language sentences recognition in first and second-language multi-talker babble. *National Center for Biotechnology Information*, *52*. 943-953.

Van Engen, K. & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America*, 121(1). 519–526.

Winet, E. D., (2012). Face-veil bans and anti-mask laws: State interest and the right to cover the face. *Hasting international and comparative law review*, *35*(1). 217

Wurm, L.H., Vakoch, D. A., Strasser, M. R., Calin,-Jageman, R., & Ross, S. E. (2001). Speech perception and vocal expression of emotion. *Cognition and Emotion*, *15*(6), 831-852.

Zinszer, B. D., Riggs, M., Reetzke, R. & Chandrasekaran B. (2019). Error patterns of native and non-native listeners´ perception of speech in noise. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.5087271

VITA

Monica A. Chulde Guayasamin was born in Quito, Ecuador. She graduated from the Central University of Ecuador in 2008 and received a Bachelor's in Sciences of Education. Upon publication of this thesis, she will have completed a M.A. in Linguistic Theory and Typology from the University of Kentucky. She currently works as an English teacher in a high school in Ecuador.