

**MONOLINGUAL AND BILINGUAL INFANTS' LANGUAGE ACQUISITIONS: CAN
INFANTS LEARN SIMILAR-SOUNDING WORDS ONLINE USING THE SWITCH
TASK**

JEAN-PHYLLIPPE PROVENCHER

Thesis submitted to the University of Ottawa
for the requirements to obtain a
Masters in clinical psychology

School of Psychology
Faculty of Social Science
University of Ottawa

© Jean-Philippe Provencher, Ottawa, Canada, 2023

Summary

When it comes to studying bilingualism and language acquisition in kids and infants, minimal pair-word learning is of particular importance due to this process being at play within a language and between languages. Having to discriminate between similar-sounding words within a language (i.e., dye & bye) and across languages (i.e., elephant & éléphant) is a complex task. The assumption that researchers have had since the early 1980s is that bilinguals may be better at certain aspects of word learning due to having to manage two languages concurrently. Previous studies have found mixed results regarding minimal pair word learning for both monolingual and bilingual groups depending on the ages. Discriminating between similar-sounding words is typically challenging before the age of 17 to 20 months regardless of the number of languages a child is learning. To investigate this question, we tested 19 participants between the ages of 14 and 18 months of age (14 monolingual and 5 bilingual). The level of attrition during the habituation trial or testing phase of the experiment made the data collection particularly difficult. Infants were recruited online from the Ottawa-Gatineau region and in Canada and the United States via a website for parents wishing to participate in various studies. During test trials, we recorded the looking time to the correctly and incorrectly labeled objects. Participants' performances were analyzed considering their language background, age, and socioeconomic status. Participants in the study were not able to discriminate between the two similar-sounding words regardless of their demographics. In conclusion, even while having access to visual cues from the speaker, monolingual and bilingual infants were not able to successfully learn similar-sounding words. Future studies should use a similar task using a live speaker via Zoom as opposed to a pre-recorded video to investigate whether the results would be different.

Acknowledgments

There are no words that can account for the support of my supervisor, Dr. Christopher Fennell. His intellect, curiosity for knowledge, and guidance have been key to my success in the program. Without your understanding and experience in academia, these last few years would have been tremendously more difficult. The recent shift in my professional trajectory would have also not been possible, and for this, I thank you sincerely.

I would also like to acknowledge my parents' support in this endeavour over the years. Despite never having had the opportunity to take part in the process of university education themselves, they have never questioned my choices and offered nothing but support and belief in what I am doing. As long as 'I am happy', it's all that matters for them...and that is the kind of support you cannot buy.

Lastly, I would like to thank my two siblings for supporting me. I have been fortunate that both my older brothers have put the educational standards low enough that surpassing my elders was possible.

Table of contents

Summary.....	ii
Acknowledgements.....	iii
Introduction	1
History of the Motor Theory of Speech Perception.....	1
Speech Perception: the role of visual gestures for word learning.....	4
History of research in bilingualism and speech perception.....	9
Hypotheses.....	14
Methods.....	14
Participants.....	14
Material.....	15
Apparatus.....	15
Procedure.....	16
Results.....	17
Socioeconomic Status.....	17
Age.....	18
Language background.....	18
Performance.....	18
Discussion.....	19
Conclusion.....	23
References.....	25
Figure	35
Appendix A: Consent form in English and French.....	36

Appendix B: Family Background Questionnaire.....39

Appendix C: Language Questionnaire.....41

Introduction

The objective of this thesis was to examine the role of relevant visual input in minimal pair-word learning in monolingual and bilingual children. Visual input in the current study refers to access to the speaker's face, with the associated rich information of mouth movements associated with particular language sounds and the engaging nature of the human face. Visual input is an integral component of speech perception and might especially play an important role when attempting to learn two similar-sounding words (e.g., boat & goat), especially when the auditory distinction is associated with different mouth movements (e.g., lip movement for "b" but not "g"). The literature on early word learning reveals that similar-sounding words have been particularly challenging for infants around the ages of 12 to 15 months and remain challenging up to 20 months. However, nearly all these studies have used recorded audio tracks with no visual input from the speaker. Furthermore, a subset of these studies investigated whether there are differences between monolingual and bilingual infants around these ages and, specifically, whether the latter group may be worse at learning these types of words due to their variable learning environment. In this document, I will first present a brief history of studies conducted with infants on word learning and speech perception, followed by studies on bilingualism and speech perception. The experiment design and methodology will then be discussed followed by the objectives and contribution of the thesis, and the hypotheses. Finally, results will be provided with the discussion and conclusion of the study.

History of the Motor Theory of Speech Perception

Researching speech perception – the process by which we hear, understand, and process sounds of languages – with high-precision technology is recent, questions regarding its process, however, are not. In 1969, the publication entitled *Perception of Speech Code* authored by

Lieberman and colleagues is still one of the most cited publications in the field several decades later (Fowler et al., 2016). The rationale of the paper was to design a machine that would read out loud, which would allow individuals who are blind to be read to via technology. Still in its infancy at the time, popular theories suggested that replacing speech sounds with a system of discrete and unique signals (speech signals) would solve the problem researchers were facing (Fowler & al., 2016). The authors challenged this view which went against previous findings that speech signals are continuous and not discrete (Fowler al., 2016).

According to Liberman's theory, speech perception is made possible by speech signals (gestures of the mouth) that are produced by the speaker but also covertly by the listener which eases the listening process. Simply put, an individual can understand and integrate perceptual information conveyed via mental representations of the other person's corresponding articulatory movements. Speech perception, according to this theory, is mirroring the production of language. This mirroring or anticipation of sound signals is believed to be a result of the brain's ability to represent these articulatory gestures with a unique phonetic pattern for each word. This process is supported to some degree by studies showing motor activation during listening tasks (Binder et al., 2004; Callan et al., 2004; Londei et al., 2010), and facilitated by mirror neurons (Kohler et al., 2002). These mirror neurons are activated during specific actions as well as throughout the observation of those same actions from another individual or species. Ferrari and colleagues (2003) found, for example, that while engaging in and observing mouth gestures, macaques had activated mirror neurons. Mirror neurons are also believed to activate the auditory system which then sets off the motor system, allowing for a representation of speech that operates automatically, meaning that an individual is not aware of this process while engaging in a conversation. Through pre-established patterns, humans are thought to be better able to engage in

this integration of speech, rapidly piecing bits of information together to provide context and meaning.

Developmentally, children have the capacity to discriminate between sounds and comprehend speech before reliably producing it. As such, it can be argued that production, although closely linked to speech perception, is not necessary to learn language. Because the motor theory of speech perception requires production for the mental representation of micro gestures, it raises immediate questions of validity and even plausibility. (Ohala, 1996; Polka & Werker, 1994). Critics that discard the motor theory of speech perception often cite and use Liberman and Mattingly's essay which contained claims and observations that held several inexact assumptions (1985). First, they claimed that speech processing is special, an argument later challenged due to its "ambiguous" meaning, and its many possible interpretations (Galantucci et al., 2006). Secondly, Liberman and Mattingly held that articulatory gestures accounted for most of speech perception; thus, minimizing the role of auditory processes and visual input, and by the same token failing to recognize the role and limits of the motor theory of speech perception. Moreover, the original theory relied solely upon association for word learning, an assumption unfounded or not exact as pre-existing cognitive structures make language learning possible (i.e., the brain has the innate structure that allows for language to emerge) (Galantucci et al., 2006). These limitations were not recognized partly due to psychology's current state at the time which was predominantly still behaviourist and the yet-to-come Cognitive Revolution that would modify theories in psycholinguistics (Miller, 2003). Indeed, language learning was perceived as an endeavour learned through experience and association alone. With the invention of computers and with it a change in understanding that the

brain is working in terms of complex wirings and structures, psychology's understanding of language shifted.

The complexity of language is still in a puzzling state even after decades of research. Contrary to critics and in response to Liberman and Mattingly's work, Moulin-Frier and Arbib (2013) argued that the Motor Theory of Speech Perception is, instead, integrated into more complex cognitive structures. Its objective, then, is not to justify speech perception in its entirety, but to offer a partial explanation of its process (Moulin-Frier & Arbib, 2013). Therefore, the role of the motor system can be viewed as one that facilitates the perception of speech. In noisy environments, for example, it is easier to understand a person facing us since articulatory gestures via visual input are not deformed by the presence of auditory noise. Such instances also stress the importance of a speaker and listener's articulatory gestures that allow the brain to better process speech and highlight the importance of visual support in the matter, a process important for this current study.

Speech Perception: *the role of visual gestures for word learning*

Despite the acknowledgment that visual input plays a role in speech perception, the extent of their implications remains uncertain (Woodhouse et al., 2009). Previous research in word learning, for example, has found reliable evidence that visual support alone provided enough information for children to acquire new words. Indeed, Havy and Zesiger (2017) assigned two conditions to children aged 30 months in which they had to learn distinct novel words: "byp", "var", "rik", "fal", "fyf", "gel", "mum", and "tit". In the audio condition, children were presented with the words without any visual input from a speaker, relying solely on auditory processes. Children were unsurprisingly able to learn novel words with audition alone. In the second condition, however, children had no access to audio and relied simply on a woman

speaker producing the words with no sound. Infants aged 30 months old were able to learn these words, demonstrating that auditory speech is not the only process through which children learn new words. Fenson and colleagues (1994) and Ganger and Brent (2004) found that babies rely more on auditory input to acquire words in the first few years of life and then later incorporate and give more attention to visual cues, particularly mouth gestures. This demonstrates the importance of visual cues in later years of life, especially in the context of word learning. Furthermore, research with babies and children has yielded evidence that children attend more to presented words when they are already able to produce these words (Yeung et al., 2013; Altwater-Mackensen & Grossmann, 2015; Guellaï et al., 2016). These findings reinforce the idea that speech perception is related, at least to a certain degree, to motor production by others and by the infant. By the same token, it points to the multimodal essence of speech integration. That is, in the same way, children can learn new words without speech, they can perceive speech with more than just visual cues or auditory sounds.

While the research mentioned points to the importance of visual input for word learning in general, this process is likely to be even more crucial for children when learning minimal pair words. Similar-sounding words or minimal pairs are words that differ solely by one phoneme. A phoneme is a unit of sound that allows us to discriminate one word from another and permit the acquisition of new words. Their differentiation implies that they convey different meanings. The words “pie” and “lie”, for example, are minimal pairs. In addition, they are also visually distinct, as they are presented differently in mouth shapes when spoken. The former starts with pursed lips and the latter with open lips and tongue protrusion. At a young age, minimal pairs are particularly challenging for infants, particularly when attempting to learn two of them at once. Werker and colleagues (2002), for instance, found that 14- month-olds confused and auditorily

presented similar-sounding words with no visual access to the speaker. By 20 months, on the other hand, participants in that study were able to learn two minimal pair words that were associated with two different objects in the habituation phase. In addition, they found that 17-month-olds also showed some ability to distinguish between similar-sounding words (some participants succeeded in the task). This is consistent with the finding that as they age, children rely more on visual speech to learn words (Havy & Zeisger, 2017).

This line of research questions in psycholinguistics stems from earlier work from Werker and colleagues (1998) who devised the *Switch task* procedure to investigate if infants were able to incorporate sound detail (i.e., close phonemes) into newly learned words. They assigned nonsense words – words not part of the vocabulary but have similar phonetic structures to native-language words – to different objects, and then, during the test trial, they violated the ascribed label to the objects to record babies' looking time to the object purposely mislabeled (Werker et al., 2002). Please see a description of the procedure below.

The Switch task is a test of infants' associative word learning, defined as the mechanism for linking a word to a referent. Researchers have asserted that perceptually-based associative word learning is the primary mode in which infants build their vocabulary (Smith et al., 2003; Sloutsky & Robinson, 2008), and others have argued that, in addition to the perceptual information necessary for associative word learning, infants also leverage conceptual information as they learn new words (i.e., their knowledge that words are shared communicative symbols; Booth et al., 2005; Waxman & Gelman, 2009). Nevertheless, both parties acknowledge and agree that associative word learning is an essential underlying mechanism in the process of early word learning. The Switch task is the most practical procedure to investigate this fundamental skill via this looking-based laboratory paradigm (Werker et al., 1998). In the

standard version of the Switch task, infants are habituated to two word-object pairings (Object A paired with Word A; Object B paired with Word B). For the procedure, both words and objects are novel to the infant. Across multiple trials, the two-word-object pairings are presented until infants' attention declines to a pre-set criterion. After this habituation, infants experience two test trials: a same trial involving one of the original word-object pairings (e.g., Object A - Word A) and a switch trial that entails a mismatched word-object pairing (e.g., Object A - Word B). If infants learned the original word-object pairings, they should detect the violation of the word-referent link during the switch trial and look longer during the switch trial than during the same trial. In another version of the task, both objects are presented simultaneously during testing trials and infants hear one of the labels. The measure then becomes the amount of looking at the object that matches the correct spoken label. In the current study, the latter version of the Switch task has been adapted online.

Werker and colleagues (1998) found that children aged 14 months and up paid more attention to the word that violated what they had previously learned, thereby showing the ability to distinguish between two very dissimilar words like "lif" and "neem" whereas younger infants could not. Subsequently, they utilized the same Switch task experiment to assess whether 14-month-olds would be able to recognize a violation of previously learned nonsense minimal pairs. Babies systematically failed to discriminate between the two similar-sounding words ("bin" & "din"). In follow-up studies, Werker and colleagues (2002) found that 20-month-olds were able to discriminate between minimal pairs, suggesting that minimal pairs are a particular category of words difficult to learn before 20 months of age. It is worth pointing out that these studies used objects as visual referents but no visual depiction of a speaker's articulatory gestures, neither onscreen nor live.

Since then, several researchers have conducted experiments to determine *why* 14-month-olds are unable to discriminate minimal pairs, and by the same token to explore “just what factors contribute to success in this task” (Fais, et al., 2012, p12.; Werker & Fennell, 2008). First, it is important to point out issues associated with the Switch task. Although widely used, the procedure remains a measure of children’s looking time at an object. Interpretations and conclusions, therefore, are indirect inferences because the learning is not explicitly stated by participants nor can it be (Fais et al., 2012). As such, these interpretations and reporting of results should be carefully assessed, especially when conducting research with bilingual children. The strength of the Switch task, however, is that it echoes the day-to-day word learning for children. At home, parents tend to point to various objects to the child and label them. A parent, for instance, will show a phone to the child and say: “Look! The phone”, which echoes the habituation phase of the Switch task.

Building on the procedure’s strength, Fais and colleagues (2012) incorporated a condition with a videotaped experimenter that presented the label words and a live experimenter presenting the spoken words, an aspect omitted by past researchers. They found that 14-month-olds failed to discriminate between minimal pair words associated with novel objects in the videotaped labeling condition. In the live experimenter condition, on the other hand, they found that children who had more mutual gaze with the experimenter recorded more looking time to the object during the switch trial. Findings that the presence of a live person is important for children to learn words bear several implications for infants and language learning. First, it reflects the significance of real-world context to acquire new words, specifically in the case of minimal pairs that are difficult for young children to learn. Second, it calls attention to early exposure to

technology and language learning for children; a growing area of research investigating the advantages and disadvantages of technology for learning language.

Considering these findings, the key to learning minimal pairs may be the presence of visual cues during the learning process. In other words, given the phonological challenge of discriminating between similar-sounding words, the articulatory cues may facilitate learning two similar-sounding words. Visual support may also be of utmost importance for bilingual children who are acquiring two languages from birth or from an early age. When bilingual children are presented with new words, there is an immediate additional challenge compared to monolinguals where they must first hear and integrate the word, but they also must hypothesize and differentiate which language the word belongs to. It leaves a multitude of avenues and possibilities for each new word presented to them, and as such, each encounter with a new word may require more than auditory information. First, to better understand the implications of the research with bilingual infants, it is worthwhile to review the history and findings in this area.

History of research in bilingualism and speech perception

The first official “studies” reported on language development in bilingual children were in the form of diaries. Parents or caregivers cataloged in detail which parent spoke what language to the child and the words and sentences children were forming. Ronjat (1913), for instance, kept records of his son learning French and German. Ronjat (1913) noted that his son did not appear to confuse the two languages in his everyday life, probably due to one parent speaking only French and the other only German. Similarly, Leopold (1970) documented in 1945, his daughter Hildegard’s progress who was also learning English and German. Again, only one parent was speaking either language. Unlike Ronjat, however, he found that Hildegard was mixing words from both languages and had grammar sentencings inconsistencies. Researchers that followed

were trying to answer a simple question stemming from these two diaries: is learning two languages at the same time detrimental for a child? In other words, will it impair or delay a child's language development? Interestingly, Leopold's claim of confusion is still transparent in today's discourse. As pointed out by Bialystok (2001), parents frequently ask her whether teaching two languages to their children will "lead to either cognitive or linguistic confusion", which is an unfounded idea (Bialystok, 2001). It is not surprising that knowledge in this area is unknown to the public given that Fennell and Lew-Williams (2017) reported that ninety percent of the literature on childhood bilingualism has been published only in the last twenty years. Bialystok (2001) further noted that this claim presupposes that language is "fragile" which is a rather curious perspective as it assumes that each language is interacting with one another, a question that is still debated among psycholinguists. Moreover, this assumption is present despite research on bilingualism having demonstrated that bilingual children reach language milestones in the same fashion that monolinguals do. (Fennell & Lew-Williams, 2017).

Hoff and Core (2015) help demystify common misconceptions for clinicians and parents by offering empirical conclusions. The following conclusions are particularly relevant considerations for this research: Dual language input does not confuse children; two languages don't need to be kept separate by parents. One parent speaking only one language does not provide better language learning than parents speaking both languages from time to time, when applicable; learning two languages takes longer than learning one which is portrayed by a lower rate for a single language development. It is not the case, however, when looking at the total vocabulary of both languages; bilingual children may exhibit strengths that are different across languages. Finally, the quantity and quality of language input help language development the most.

Researchers working with monolingual and bilingual infants have found that in the first 12 months of life, infants show sensitivity toward their native languages (Bahrick & Pickens, 1988; Dehaene-Lambertz & Houston, 1998; Mehler et al., 1988; Moon et al., 1993). Shortly after birth, researchers have found that infants can perceive language differences based on prosody, or rhythm (Mehler & Christophe, 1995; Mehler et al., 1998). For example, French only places stress at the end of an entire phrase, whereas English is a stress-timed language that has a strong-weak pattern where they often stress the first syllable of each word. Thus, newborns can discriminate French from English because of their rhythmic differences. When languages are from the same prosodic family, however, newborns are not able to reliably discriminate between the two. An interesting development case is when infants are learning two prosodically similar languages. Bosch and Sebastian-Gallés (2001) studied 4-month-old bilingual infants to determine if slightly older infants could discriminate between Spanish and Catalan which have similar prosody. In their study, researchers found that bilingual infants were able to show some ability to discriminate between languages with similar rhythms by 4.5 months. Results therefore suggest already after just a few months bilingual infants can already discriminate between native languages with similar prosodies (Bosch & Sebastian-Gallés, 2001).

Other researchers investigating phonemes with infants have found that while infants refine their discrimination of phonemes from a broad universal set to those in their native language(s), this refinement typically develops initially with vowels and eventually applies to consonants (Sundara et al., 2018). Bosch and Sebastian-Gallés (2003) wanted to explore this phonological refinement in bilingual infants. In their studies, they first analyzed if 4- and 8-month-old Spanish-Catalan bilingual and Catalan monolingual infants were able to perceive vowel contrasts that exist in Catalan and not Spanish, specifically testing the contrast /e/ versus

/ε/. In this first experiment, both 4-month-old Catalan monolinguals and Spanish-Catalan bilingual participants succeeded in perceiving the difference regardless of their native language (Bosch & Sebastian-Gallés, 2003). At 8 months of age, however, only monolingual Catalan participants succeeded in discriminating the two sounds, leading researchers to posit that mere language exposure does not dictate the ability of infants to discriminate close vowel contrasts. In their last experiment, they found that after their first year of life (12 months), bilingual infants were once again able to discriminate between the two sounds. This finding is particularly important for our current research given that infants at this age can discriminate between sounds that differ only by phoneme contrast (same letter sound). Although minimal pairs are not simple phoneme contrasts, they also differ only by one phoneme (i.e., pie & bye).

Building on their previous research's findings that monolingual infants are not able to learn the similar-sounding words /bih/ and /dih/ before 17 months, Fennell and colleagues (2007) found that bilinguals were not able to successfully learn similar-sounding words until they reached 20 months of age with a native contrast. Although there may be a myriad of explanations, researchers have posited that bilinguals' need to manage two languages and having to distinguish between similar-sounding words which are difficult may lead to a delay in their ability to display the same skills compared to monolinguals. In a more recent study, researchers investigated whether bilingual Spanish-English-learning infants of 14 and 19 months can learn two words using the same syllable but with varying pitch sounds – mimicking the lexical differences seen in languages like Mandarin (Graf Estes & Hay, 2015). Using the syllables /ku/ (falling sound) and /ku/ (rising sound), bilingual participants were able to detect when the label and the object pairing did not match, demonstrating that they have learned the words and are capable of distinguishing between pitch contours. This is important given that our study will be

investigating infants between the ages of 14 and 18 months who will have to learn similar-sounding words. In a follow-up experiment with 22-month-old bilinguals, participants were no longer able to learn pitch contour. The authors posited that as children become more experienced with language, they have less flexibility when it comes to considering pitch contours as two different labels (Graf Estes & Hay, 2015).

Pons and colleagues (2015) investigated the importance of visual cues for infants. They found that monolingual infants pay more attention to the eyes when learning native and non-native words (4 months old), look longer at the mouth at 8 months, and look longer at the mouth at 12 months of age for nonnative words only. This demonstrates that by 12 months of age, monolingual infants only paid attention to the mouth for sounds they were not familiar with. Bilingual infants speaking Catalan and Spanish, at the age of 4 months paid attention to both the eyes and the mouth. At 8 months, bilingual infants spend more time looking at the mouth, and by 12 months they pay even more attention to the mouth for native and nonnative words. This suggests that visual cues play a significant role in bilingual infants learning and integrating new words into their vocabulary. Due to bilingual infants always having access to more than one language, they may have to spend more time using visual cues to distinguish which language is being spoken to extrapolate meaning.

Hypothesis

Using the Switch task procedure, the purpose of this study was to investigate the role visual access to speech plays in minimal pair word learning. In other words, does having visual access to a speaker's mouth gestures facilitate the learning of similar-sounding words? We hypothesize that monolingual infants will be better able to perceive minimal pair changes due to having to only deal with one language as opposed to two. It is possible, on the other hand, that

bilingual infants may be better at perceiving these changes due to their overall attentional advantage, stemming from having to constantly attend to phonological language differences. Previous studies have shown that bilingual children spend more time looking at a speaker's mouth at the age of 8 months than monolinguals (Ayneto & Gallés, 2017). Meanwhile, however, they also found that at the age of 12 months, bilinguals and monolinguals paid equal attention to the mouth. Although this study will not look at where children are paying attention, we can deduce that having visual access will give children a significant chance of potentially being able to learn minimal pairs.

Methods

Participants

For the study, the final sample collected was comprised of 19 participants (10 males and 9 females) typically developing children aged between 14 and 18 months of age. Exclusion criteria included children with developmental delays, including hearing and vision difficulties due to task demands. Fourteen participants were monolingual and 5 were bilingual. To be considered bilingual, participants recorded at least 20% exposure to a second language weekly. After parents reported speaking more than one language at home (or at daycare), participants' language exposure was assessed using an online adaptation of the Language Exposure Questionnaire (Bosch & Sebastian-Gallés, 1997).

Research has shown that by the age of 15 months, monolingual children are capable of learning words, with fidelity, in the laboratory (Fennell & Waxman, 2010; Byers-Heinlein et al., 2013). By the age of 20 months, bilingual children can discriminate newly learned words that differ by one phoneme (Fennell et al., 2007). Although monolingual and bilingual children can learn words in their natural environment by the age of six months, it is not before 12 to 14 months that they can reliably do so in an experimental context (Bergelson & Swingley, 2012).

Participants were recruited from the Ottawa-Gatineau region, and some were recruited via the <https://childrenhelpingscience.com> website. Parents submitted a request to participate in the study and were contacted to set up a time to participate. The location of families varied, with most participants living in Canada and others living in the United States.

In this study, there was a significant level of participant exclusion ($n = 64$) due to failure to complete the task. Infants who failed to complete the task either stopped paying attention, became too bored, or were unable to tolerate the attention required to complete the task (roughly 10 minutes total for the habituation phase and testing trials).

Material

The stimulus words used in this study were the novel words “bos” and “gos”. The two words were given a referent image that is also novel for infants, a green stapler (gos) and a blue funnel (bos).

Apparatus

The habituation phase and testing phase were recorded using the platform Zoom to allow researchers to code looking time post-testing. The looking time for the correct object during testing trials was coded using the software ELAN (<https://sites.bu.edu/elsa/elan-coding/>).

Procedure

The data was collected by the primary researcher. Participants were all tested using the online adaptation of the Switch task. Prior to the Zoom call, parents signed the consent form, filled out the Language Exposure Questionnaire if applicable to their child, and filled out a demographic questionnaire. Upon joining the Zoom, parents were welcomed, and the specifics of the experiment were discussed with parents (i.e., it is preferable to have their child on their lap, remove any distractions around their space, etc.). The researcher confirmed once again that

parents wished for the child to participate in the study. Each parent was explained that the experiment could stop at any time should their child experience discomfort, boredom, or any other negative emotions.

The experiment had two phases that every participant engaged in, the habituation trial and the testing phase.

Habituation phase

During the habituation phase, the words “bos” and “gos” were presented in discrete trials by the speaker and each associated with a distinct object. The words were chosen for two reasons: linguistically, they are both English and French-appropriate, and the words were used in previous studies. Although the purpose of this study was not to compare with other studies, gathering data using the same words could be used by later researchers. The word “bos” was presented with a blue funnel, and “gos” was associated with a green stapler. On-screen, the female speaker appeared on the right side of the screen and the objects were shown on the left side. The speaker’s speech articulatory movements were always visible during the habituation phase. Each word was presented 10 times per trial. The maximum amount of habituation or learning trials participants could have was 24. During one learning or habituation trial, the word being said by the female speaker is done repeatedly with different voice tones using infant-directed speech (e.g., Bos, bos, bos, bos, bos). The purpose of the habituation trial was twofold: to ensure that children can pay attention and learn the words they are presented with, and at the same time to familiarize themselves with the upcoming testing phase. The habituation phase lasts until the infant has reduced their looking to the screen by 50% (measured in blocks of 4 trials) or when the infant reaches a maximum of 24 trials. Most participants completed the 24 trials, rather than reaching the habituation criterion of reduced looking time. This is quite unusual in Switch

task experiments. It is likely due to the task being compelling enough with a visual speaker that they did not reduce their looking time by half.

Testing phase

During the testing phase, participants are shown the previously presented words “bos” and “gos”. Unlike the habituation phase, the speaker was not present on the screen. Although having a speaker during the testing phase could be investigated, we determined that it would provide more visual complexity for children, leading them to look more to the person on the screen than the objects. Bilingual and monolingual children saw both objects simultaneously on the screen and heard either “bos” or “gos” in various sequences during 8 testing trials. Similar to the habituation phase, participants heard the given word several times during one testing trial. Using the ELAN coding software, looking time for the correct object was coded.

Results

Socioeconomic Status

Out of the 19 participants, the parents of three infants omitted the demographic questionnaire. While the language status of the children is known based on intake data (whether they are monolingual or bilingual), the economic status and highest diploma obtained by the families were not known. The value attributed to each diploma was as follows: 1 = High School / GED, 2 = College or Trade, 3 = Bachelor, 4 = Master, 5 = Medical degree, Jurisprudence degree or Doctor of Dental Surgery, 6 = Doctorate. The average for both bilingual and monolingual parents was a Master. As such, missing data were assigned a 4. There was no statistical significance for education between participants.

Age

The average age of participants was initially calculated with three missing ages (two for monolingual participants and one for bilingual participants). The average age for monolingual was 14.9 whereas the average age for bilingual participants was 15. The final age average age of participants was 14.94. There was no statistical significance in age between groups or during performance.

Language Background

The percentage of language exposure for participants was calculated based on information provided by parents on the number of hours each parent or caregiver spent with the participant and the languages they speak with them during those hours. The percentage was then calculated based on the number of hours participants were awake and the percentage they heard one language versus the other. The language exposure percentage varied for participants between 43% and 57% for most bilingual participants ($n=3$) and one participant had less exposure to her second language 21% (lowest) and 77% (highest). It is worth noting that the latter had exposure to a third language, however, her exposure to English was less than 5% and therefore the calculations do not add up to 100%. There was no statistical significance of performance for the group (see below).

Performance

An exploratory analysis to determine if there was a gender effect. We performed a mixed analysis of variance (2x2 ANOVA) with 2 factors (test trial; correct vs incorrect) x (gender; male vs female). The interaction between the test trials and gender was not significant, $F(1,17) = .01$, $p = .93$. However, there was a between-subject main effect of gender $F(1,17) = 5.03$, $p = .04$. Despite both genders failing to learn minimal pairs by showing no difference in reaction to correct or incorrect stimuli, $F(1,17)$, 1.43, $p = .25$, males were looking longer than females

overall during the testing phase by 6.36 seconds to the correctly labeled object, but also by 6.78 seconds to the incorrectly labeled object.

We then analyzed the interaction, if any, between monolinguals and bilinguals and their performance on the task. There was no interaction between correct versus incorrect across groups, $F(1.17), 1.43, p=.25$ and no main effect of group, $F(1.17), 1.31, p=.27$, meaning that both monolingual and bilingual participants were not reacting differently across test trials (see Figure 1 for looking time performance). However, we acknowledge that this analysis is problematic given the disparity between group sample sizes.

Although there was no effect of group, we conducted a posthoc analysis with monolingual participants only due to the imbalance between the number of bilingual and monolingual infants to investigate any possible difference in pattern. The results of the paired t-test showed no difference between the amount of time monolingual infants looked to the correct versus incorrect object, $t(13) = .83, p = .42$ (see Figure 1 for looking time performance).

Discussion

The objective of the thesis was to investigate whether monolingual and bilingual infants between the ages of 14 and 18 months can discriminate between two similar-sounding words when provided access to visual cues (a speaker's facial expressions and mouth gestures). Historically, researchers have found that before 20 months of age, infants cannot reliably learn two new objects when they differ by one phoneme without additional learning support (Werker & Fennell, 2002).

We hypothesized that monolinguals may be better at discriminating minimal pairs given that they only have to manage one language. It was also possible, on the other hand, that bilinguals would be better given that they may pay more attention to visual cues from the

speaker's mouth than their monolingual peers. Additionally, given that they hear two languages in their day-to-day life, they may be better at perceptual discrimination. Results indicated that neither monolingual nor bilingual infants were successfully able to discriminate between similar-sounding words, a result consistent with previous findings that before 20 months. Infants have difficulties learning minimal pairs, around 17 months they show some ability, and by 20 months of age they can reliably learn minimal pairs (Werker et al., 2002). Results are concurrent with 18 out of the 19 participants' age falling below 17 months. It is worth noting results should be interpreted with caution due to the level of attrition during the habituation trial and testing phase of the experiment, and the small sample size of bilingual participants. Recruiting bilingual infants is typically a challenging task compared to monolinguals. In this study, it was even more difficult due to the attrition and relatively small number of bilingual infants that we were able to recruit, especially during the pandemic. With a larger sample, it is possible that bilingual infants may perform differently as a group. It is also possible that infants slightly older may not perform better unless the task, when using Zoom, is done with a live experimenter- that is providing social contingency. Indeed, Fais and colleagues (2012) did find that infants failed in the videotaped condition but were able to discriminate between minimal pairs as early as 14 months of age when engaging in mutual gaze with the experimenter. It is worth noting that pre-recorded stimuli are simpler to ensure that each participant receives the same habituation training and testing trials. It is possible, however, that minimal pairs are too difficult to learn at this age unless they are mitigated via social interaction.

Although there is a range in representation regarding the sociodemographic status (SES) of families that participated in our study, the average family was of a higher economic status (making over 100,000\$ a year in the family unit) and of higher education with most parents

having a bachelor's degree, a master's degree, or a Ph.D. Despite the lack of representation, typically higher education and SES are linked to increased cognitive performance, specifically when it comes to language and language exposure, and executive functions (Hackman & Farah, 2008). Given these research findings, infants who participated in the study had, if anything, a family background, and economic situation that would have been favourable for them to succeed in this task. Despite the possible advantages stemming from higher SES, our results showed no differences in performance, with both monolingual and bilingual groups not detecting a switch in labels.

Given that both groups are failing to discriminate minimal pairs, it can also be posited that monolingual and bilingual infants are following a similar developmental trajectory when it comes to this specific area of word learning. At this age, infants are in a developmental stage where they are beginning to associate previously known and new words with objects at an increasing rate (Werker, et al., 2002). When it comes to learning to associate similar-sounding words to newly presented objects, however, the task remains too challenging for 14 to 16-month-old infants. Although many explanations are possible, researchers have put forward that linking words to objects is a difficult task in and of itself due to the computational demand it requires (Stager & Werker 1997). Infants need to be able to attend and sustain attention to learn two new words that differ by only one phoneme in a short period of time, and they need to associate their meanings with the newly presented objects. During the test trials, infants must detect a disruption in their learning (the label switch) which differs by a small phonetic detail, all the while having to sustain attention throughout this process. Around the second year of life, typically infants' vocabulary rapidly increases with words that may be similar in phonetic detail (i.e., bee & bye) thus requiring this ability to integrate more words to navigate their environment and by extension

increasing their vocabulary. By the time they reach 20 months of age typically known as language spurt, infants may have reached a developmental stage where they have mastery over contrasting phonetically similar words. Because there is increased vocabulary around the second year of life, there is pressure for infants to have to develop this skill. This could also reinforce one of the assumptions of the motor theory laid out by Liberman that speech perception is mirroring the production of language. In other words, infants can produce more words due to visual perception and vice-versa, making them better at discriminating between similar-sounding words which is a task that eventually becomes effortless throughout childhood and in adulthood.

As the results showed, female participants looked less in terms of total time than their male counterparts by 6.37 seconds for the correct object and 6.78 seconds for the incorrect object. Although this can be interpreted in many ways, our hypothesis is that it is possible that females process the stimuli to a greater degree during the habituation trial (learning phase) and therefore are less interested in attending to the stimuli at test. This hypothesis is based on the enhanced language development of females over males early in life. Researchers have found that from birth to slightly over 24 months, girls are better at using communicative gestures, produce more vocabulary and are slightly better at combining words (Eriksson et al., 2012).

It is worth noting that despite the Switch task being one of the most common methods to assess novel word learning at these ages, we conducted our study in a novel manner due to COVID-19 restrictions: an online version of the task. Using Zoom to determine if infants are able to learn words is a methodology that, to my knowledge based on a literature review, has rarely been used, and never for minimal pair learning. In 2014, Roseberry, Hirsh-Pasek and Golinkoff investigated social contingency when it comes to word learning at 24 to 30 months of age. In their study, infants were only able to learn dissimilar sounding novel verbs having

socially contingent interactions (i.e., “Look at what I can do with this toy! I'm /meeping/ it!”) in a live manner. Thus, again, it appears that social contingencies may be important to facilitate word learning when using videotaped speakers. This is interesting because infants do not need such social contingencies when presented with word-object pairings without humans on screen. There is a vast literature, including nearly all the Switch task literature, wherein infants successfully learn words via spatial-temporal contingencies. Essentially, most successful word learning studies present an object (or action) on-screen contiguous with an audio presentation of the target word. It would be interesting to investigate whether the presence of a human speaker on screen drives a different set of word learning expectations, where simple spatial-temporal association is no longer sufficient.

Throughout the time we tested participants, various patterns of attrition emerged due to infants' inability to complete all testing trials, primarily becoming fussy (or crying) due to the attentional demand of the task. Although attrition is common with infants using the Switch task, the rates appeared to be different across time and were much higher than in past studies. For instance, early in the pandemic, participants' attrition was similar to other Switch studies with infants (50-60% attrition rate) while in the later stages of the pandemic and after restrictions were lifted, attrition spiked between 80-90% even though no changes in methodology occurred. It is possible that early during the pandemic, participants were used to staying in-door and attending to the same activity for longer periods of time whereas later in the pandemic infants were now used to changing activities and environment (going to another room) more often.

It is also noteworthy that having access to a recorded speaker appears may have prevented infants from habituating. The speaker may have been so interesting to the infants that they maintained attention throughout the learning phase and did not habituate. This lengthened

the habituation phase and would cause problems for the infants in terms of their overall ability to stay engaged during the latter parts of the experiment (i.e., the test phase). On the rare occasions that participants habituated, they did so due to not being able to engage with the task anymore and failing to complete the experiment. These participants had to be removed due to insufficient looking time to either object (no longer paying attention). Future studies should investigate attrition in the context of word learning using an online version of the Switch task.

Conclusion

Our study aimed to investigate whether monolingual and bilingual infants can learn minimal pairs having access to visual cues, specifically a speaker's facial expression and mouth gestures. The Switch task was used in an online format via Zoom, a new way of investigating word learning that departs from traditional laboratory testing. Contrary to our hypotheses, both monolingual and bilingual infants failed to learn minimal pairs even when provided visual cues. Results reaffirm previous findings that before at least 17 months of age, infants are not able to distinguish similar-sounding words simultaneously. Once again, these results are asking an important question: why is learning minimal pairs so challenging for infants and what is the mechanism that allows them to succeed shortly thereafter?

References

- Altwater-Mackensen, N., & Grossmann, T. (2015). Learning to match auditory and visual speech cues: Social influences on acquisition of phonological categories. *Child development, 86*(2),362-378. <https://doiorg.proxy.bib.uottawa.ca/10.1111/cdev.12320>
- Aslin, R. N. (2007). What's in a look?. *Developmental science, 10*(1), 48-53. <https://doi-org.proxy.bib.uottawa.ca/10.1111/j.1467-7687.2007.00563.x>
- Ayneto, A., & Sebastian-Galles, N. (2017). The influence of bilingualism on the preference for the mouth region of dynamic faces. *Developmental science, 20*(1). <https://doi-org.proxy.bib.uottawa.ca/10.1111/desc.12446>
- Badino, L., D'Ausilio, A., Fadiga, L., & Metta, G. (2014). Computational validation of the motor contribution to speech perception. *Topics in cognitive science, 6*(3), 461-475. <https://doi-org.proxy.bib.uottawa.ca/10.1111/tops.12095>
- Bahrack, L. E., & Pickens, J. N. (1988). Classification of bimodal English and Spanish language passages by infants. *Infant Behavior and Development, 11*(3), 277-296. [https://doi.org/10.1016/0163-6383\(88\)90014-8](https://doi.org/10.1016/0163-6383(88)90014-8)
- Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253-3258. <https://doi-org.proxy.bib.uottawa.ca/10.1073/pnas.111338010>
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press. <https://doi-org.proxy.bib.uottawa.ca/10.1080/15235882.2002.10162587>

- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature neuroscience*, 7(3), 295-301. <http://dx.doi.org.proxy.bib.uottawa.ca/10.1038/nn1198>
- Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, 65(1), 33-69. [https://doi.org/10.1016/S0010-0277\(97\)00040-1](https://doi.org/10.1016/S0010-0277(97)00040-1)
- Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, 2(1), 29-49. https://doi-org.proxy.bib.uottawa.ca/10.1207/S15327078IN0201_3
- Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and speech*, 46(2-3), 217-243. <https://doi-org.proxy.bib.uottawa.ca/10.1177/0023830903046002080>
- Booth, A. E., Waxman, S. R., & Huang, Y. T. (2005). Conceptual information permeates word learning in infancy. *Developmental psychology*, 41(3), 491. <https://doi-org.proxy.bib.uottawa.ca/10.1037/0012-1649.41.3.491>
- Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences*, 112(44), 13531-13536. <https://doi-org.proxy.bib.uottawa.ca/10.1073/pnas.1508631112>
- Byers-Heinlein, K., Fennell, C. T., & Werker, J. F. (2013). The development of associative word learning in monolingual and bilingual infants. *Bilingualism: Language and Cognition*, 16(1), 198-205. <https://doi-org.proxy.bib.uottawa.ca/10.1017/S1366728912000417>

- Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native-and second-language speakers differentially activates brain regions involved with acoustic-phonetic processing and those involved with articulatory–auditory/orosensory internal models. *NeuroImage*, 22(3), 1182-1194.
<https://doi.org/10.1016/j.neuroimage.2004.03.006>
- Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. F. (2017). The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive Development*. <https://doi.org/10.1016/j.cogdev.2017.02.004>
- Dehaene-Lambertz, G., & Houston, D. (1998). Language discrimination response latencies in two-month-old infants. *Language and Speech*, 41(1), 21-43.
<https://doi-org.proxy.bib.uottawa.ca/10.1177/002383099804100102>
- Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., ... & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British journal of developmental psychology*, 30(2), 326-343. <https://doi.org/10.1111/j.2044-835x.2011.02042.x>
- Fais, L., Werker, J. F., Cass, B., Leibowich, J., Barbosa, A. V., & Vatikiotis-Bateson, E. (2012). Here's looking at you, baby: what gaze and movement reveal about minimal pair word-object association at 14 months. *Laboratory Phonology*, 3(1), 91-124.
<https://doi.org/10.1515/lp-2012-0007>
- Fennell, C. T., Byers-Heinlein, K., & Werker, J. F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child development*, 78(5), 1510-1525.
<https://doi.org/10.1111/j.1467-8624.2007.01080.x>

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child development*, *81*(5), 1376-1383.

<https://doi.org/10.1111/j.1467-8624.2010.01479.x>

Fennell, C., & Lew-Williams, C. (2017). Early bilingual word learning. *Early word learning*, 110-122. <https://doi-org.proxy.bib.uottawa.ca/10.1098/rstb.2009.0105>

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... & Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, i-185. <https://doi.org/10.2307/1166093>

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, *17*(8), 1703-1714.

<https://doi.org/10.1046/j.1460-9568.2003.02601.x>

Fowler, C. A., Shankweiler, D., & Studdert-Kennedy, M. (2016). " Perception of the speech code" revisited: Speech is alphabetic after all. *Psychological review*, *123*(2), 125.

<https://doi.org/10.1037/rev0000013>

Fowler, C. A. (2015). COSMO's "motor theory" is not the motor theory of Liberman, Cooper, and Mattingly. *Journal of Phonetics*, *53*, 42-45.

<https://doi.org/10.1016/j.wocn.2015.06.002>

Fowler, C. A., & Galantucci, B. (2005). The relation of speech perception and speech production. *The handbook of speech perception*, 632-652.

<https://doi.org/10.1002/9780470757024>

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, *13*(3), 361-377.

<https://doi-org.proxy.bib.uottawa.ca/10.3758/BF03193857>

- Ganger, J., & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental psychology*, 40(4), 621. <https://doi.org/10.1037/0012-1649.40.4.621>
- Guellai, B., Streri, A., Chopin, A., Rider, D., & Kitamura, C. (2016). Newborns' sensitivity to the visual aspects of infant-directed speech: Evidence from point-line displays of talking faces. *Journal of Experimental Psychology: Human Perception and Performance*, 42(9), 1275. <https://doi.org/10.1037/xhp0000208>
- Grabski, K., Tremblay, P., Gracco, V. L., Girin, L., & Sato, M. (2013). A mediating role of the auditory dorsal pathway in selective adaptation to speech: a state-dependent transcranial magnetic stimulation study. *Brain research*, 1515, 55-65. <https://doi.org/10.1016/j.brainres.2013.03.024>
- Graf Estes, K., & Hay, J. F. (2015). Flexibility in bilingual infants' word learning. *Child development*, 86(5), 1371-1385. <https://doi.org/10.1111/cdev.12392>
- Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in cognitive sciences*, 13(2), 65-73. <https://doi.org/10.1016/j.tics.2008.11.003>
- Havy, M., & Zesiger, P. E. (2017). Learning Spoken Words via the Ears and Eyes: Evidence from 30-Month-Old Children. *Frontiers in Psychology*, 8, 2122. <https://doi.org/10.3389/fpsyg.2017.02122>
- Ivry, R. B., & Justus, T. C. (2001). A neural instantiation of the motor theory of speech perception. *Trends in Neurosciences*, 24(9), 513-515. [https://doi.org/10.1016/S0166-2236\(00\)01897-X](https://doi.org/10.1016/S0166-2236(00)01897-X)
- Kohler, E., Keysers, C., Umiltà, M., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science (New*

- York, N.Y.), 297(5582), 846-8. <https://doi-org.proxy.bib.uottawa.ca/10.1126/science.1070311>
- Leopold, W. F. (1970). *Speech Development of a Bilingual Child: Grammar and general problems in the first two years*(Vol. 3). Ams Press.
- Lewkowicz, D. J., & Lickliter, R. (Eds.). (2013). *The development of intersensory perception: Comparative perspectives*. Psychology Press.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1-36. [https://doi.org/10.1016/0010-0277\(85\)90021-6](https://doi.org/10.1016/0010-0277(85)90021-6)
- Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Gratta, C. D., Romani, G. L., & Belardinelli, M. O. (2010). Sensory-motor brain network connectivity for speech comprehension. *Human brain mapping*, 31(4), 567-580. <https://doi.org/10.1002/hbm.20888>
- Hoff, E., & Core, C. (2015, May). What clinicians need to know about bilingual development. In *Seminars in speech and language* (Vol. 36, No. 2, p. 89). NIH Public Access. <https://doi-org.proxy.bib.uottawa.ca/10.1055/s-0035-1549104>
- Holt, L. L., & Lotto, A. J. (2014). The alluring but misleading analogy between mirror neurons and the motor theory of speech. *Behavioral and Brain Sciences*, 37(2), 204-205. <https://doi-org.proxy.bib.uottawa.ca/10.1017/S0140525X13002331>
- Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisited. *Psychonomic bulletin & review*, 15(2), 453-457. <https://doi.org/10.3758/PBR.15.2.453>

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2), 143-178.

[https://doi.org/10.1016/0010-0277\(88\)90035-2](https://doi.org/10.1016/0010-0277(88)90035-2)

Mehler, J., & Christophe, A. (1995). Maturation and learning of language in the first year of life.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in cognitive sciences*, 7(3), 141-144. [https://doi.org/10.1016/S1364-6613\(03\)00029-9](https://doi.org/10.1016/S1364-6613(03)00029-9)

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant behavior and development*, 16(4), 495-500.

[https://doi.org/10.1016/0163-6383\(93\)80007-U](https://doi.org/10.1016/0163-6383(93)80007-U)

Moulin-Frier, C., & Arbib, M. A. (2013). Recognizing speech in a novel accent: the motor theory of speech perception reframed. *Biological cybernetics*, 107(4), 421-447.

<https://doi.org/10.1007/s00422-013-0557-3>

Ohala, John J. (1996). Speech Perception Is Hearing Sounds, Not Tongues. *Journal of the Acoustical Society of America*, 99(3), 1718-25. <https://doi.org/10.1121/1.414696>

Plant, G. (1999). Perceiving Talking Faces: From Speech Perception to a Behavioral Principle. *Ear & Hearing*, 20(2), 178-179.

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological science*, 26(4), 490-498.

<https://doi.org/10.1177/0956797614568320>

Ronjat, J. (1913). *Le développement du langage observé chez un enfant bilingue*. H. Champion.

Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child development*, 85(3), 956-970.

<https://doi.org/10.1111/cdev.12166>

- Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain & Language*, 111, 1–7.
<https://doi.org/10.1016/j.bandl.2009.03.002>
- Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action-candidate roles for motor cortex in speech perception. *Nature reviews. Neuroscience*, 10(4), 295. <https://doi.org/10.1038/nrn2603>
- Singh, L., Fu, C. S., Tay, Z. W., & Golinkoff, R. M. (2017). Novel Word Learning in Bilingual and Monolingual Infants: Evidence for a Bilingual Advantage. *Child Development*.
<https://doi.org/10.1111/cdev.12747>
- Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in infants' visual processing: From overshadowing to attentional tuning. *Cognitive Science*, 32(2), 342-365.
<https://doi.org/10.1080/03640210701863495>
- Smith, L. B., Jones, S. S., Yoshida, H., & Colunga, E. (2003). Whose DAM account? Attentional learning explains Booth and Waxman. *Cognition*, 87(3), 209-213.
[https://doi.org/10.1016/s0010-0277\(02\)00236-6](https://doi.org/10.1016/s0010-0277(02)00236-6)
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388(6640), 381-382. <https://doi.org/10.1038/41102>
- Stasenko, A., Bonn, C., Teghipco, A., Garcea, F. E., Sweet, C., Dombovy, M., ... & Mahon, B. Z. (2015). A causal test of the motor theory of speech perception: A case of impaired speech production and spared speech perception. *Cognitive neuropsychology*, 32(2), 38-57. <https://doi.org/10.1080/02643294.2015.1035702>

- Sundara, M., Ngon, C., Skoruppa, K., Feldman, N. H., Onario, G. M., Morgan, J. L., & Peperkamp, S. (2018). Young infants' discrimination of subtle phonetic contrasts. *Cognition*, *178*, 57-66. <https://doi.org/10.1016/j.cognition.2018.05.009>
- Turner, A. C., McIntosh, D. N., & Moody, E. J. (2015). Don't listen with your mouth full: the role of facial motor action in visual speech perception. *Language and speech*, *58*(2), 267-278. <https://doi.org/10.1177/0023830914542305>
- Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in cognitive sciences*, *13*(6), 258-263. <https://doi.org/10.1016/j.tics.2009.03.006>
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental psychology*, *34*(6), 1289. <https://doi.org/10.1037/0012-1649.34.6.1289>
- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1-30. https://doi.org/10.1207/S15327078IN0301_1
- Werker, Janet F. & Christopher T. Fennell (2008). Infant speech perception and later language acquisition: Methodological underpinnings. In John Colombo & Peggy McCardle (eds.), *The measurement of language in infancy*, 85–98. Hillsdale, NJ: Lawrence Erlbaum Associates. ^[1]_[SEP]
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*, 701–702. <https://doi.org/10.1038/nn1263>

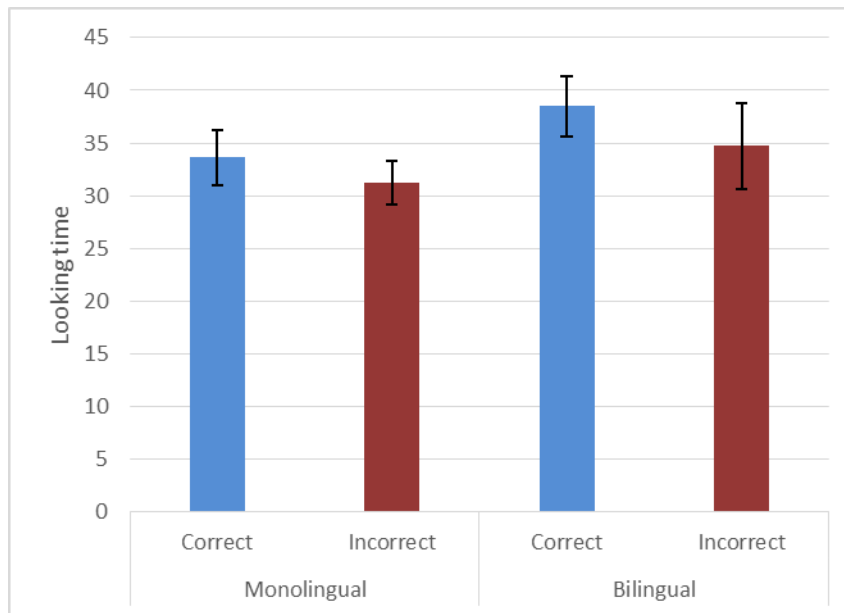
Wilson, S. M. (2009). Speech perception when the motor system is compromised. *Trends in cognitive sciences*, 13(8), 329. <https://doi.org/10.1016%2Fj.tics.2009.06.001>

Woodhouse, L., Hickson, L., & Dodd, B. (2009). Review of visual speech perception by hearing and hearing-impaired people: clinical implications. *International Journal of Language & Communication Disorders*, 44(3), 253-270. <https://doi.org/10.1080/13682820802090281>

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68(2), 123-139. <https://doi.org/10.1016/j.jml.2012.09.004>

Figure 1.

Total Looking Time to the Correct and Incorrect Object per Group. Error bars reflect the standard error of the mean.



Appendix A: Consent Forms



Université d'Ottawa | University of Ottawa

École de psychologie | School of Psychology

136 Jean-Jacques Lussier, Vanier 3002, Ottawa, ON K1N 6N5

☎ 613-562-5801 📠 613-562-5147

Consent Form: Language Development Lab Monolingual and Bilingual Infants' Attention to and Use of Visual Cues in Word Learning

Principal Investigator: Dr. Christopher Fennell, University of Ottawa, Faculty of Social Sciences, School of Psychology, (613) 562-5800 Ext. 4445, fennell@uottawa.ca.

Purpose: We are interested in understanding the development of monolingual and bilingual infants' attention to and use of visual cues from talkers' mouths to learn similar-sounding words.

Study Procedure: The experiment will take place in the comfort of your home using your computer. We request that the experiment take place in a quiet room with minimal distraction. It is a maximum of nine minutes long and set-up for the experiment will take approximately two minutes. Your infant will either be seated on your (the parent's) lap or in a highchair in front of the computer screen, your choice. Your infant will see a woman sitting at a table looking straight ahead and saying one nonsense word. While repeating the word several times, the woman will present an object to be associated with the word. Then, she will say a new, but similar-sounding, word while a new object is presented. Depending on what experimental condition your baby is in, the woman's mouth will either be shown or blocked from view. If the mouth is shown, the audio may be turned off to see if babies can learn words from visual cues alone. These trials will repeat until your baby becomes bored, which is a sign of learning. We will then show both objects on the screen and your baby will hear one of the two object labels. Looking at the labelled object is our measure of learning.

Your infant's reactions to the stimuli, specifically his or her looking time and location to the screen, will be video recorded and coded by the researcher. We will ask you to look away the entire time once the study begins so that your reactions to the stimuli do not interfere with the study.

We will ask you (the parent) to fill out a demographic questionnaire (sent via email ahead of time) and will also ask you a few questions about the language(s) that your infant is learning after the experiment via Zoom.

Risks: Participation in this study does not involve any risks or discomfort to you or your infant.

Benefits: There is no direct benefit to your infant by his/her participation. A possible benefit for parents is gaining a greater understanding of infant language development. This research will also benefit society by giving researchers and educators a better understanding of language development in monolingual and bilingual infants.

Participant’s Rights: You and your infant’s participation in this study is completely voluntary and you or your infant are free to withdraw at any time. You may at any time refuse to answer questions if you are uncomfortable doing so. If your baby becomes upset during testing, we will stop the study for your baby’s comfort. Participation or withdrawal will not affect any rights to which you are entitled. Withdrawal will in no way affect any present or future involvement with the University of Ottawa or its services. If you choose to withdraw from the study, your data will be destroyed, unless permission to use it is granted.

As part of our public and inclusive mandate at the uOttawa Living Lab our studies are being conducted with a wide range of participants (ages, language backgrounds etc.). I understand that my individual child’s data may not be immediately used for the current research data analyses; however, they may be used later as part of larger analyses of development and may inspire future research. Where applicable, ethics approval for secondary use of data will be obtained.

Confidentiality and conservation of data: Any information resulting from this research will be kept strictly confidential. Participants will not be identified by name in any reports of the completed study. All electronic files and hard copies of the data (e.g., printouts) will be identified only by code and kept in a secure location. Only members of the lab will have access to the data. Hard copies of the data will be conserved for a minimum of 10 years. We will ask you whether you agree to having the researchers use segments of the video records for educational (e.g., classroom instruction) and academic (e.g., conference presentations) purposes. We will not reveal your infant’s name in these presentations. Please note that faces will not be blurred in these videos. Lastly, by agreeing to participate in the study, participants are aware and acknowledge the potential confidentiality risks associated with using an online platform such as Zoom.

Compensation: There is no compensation for this study.

Contact: If you have any questions or desire further information with respect to this study, you may contact **Dr. Christopher Fennell**. If you have any questions regarding the ethical conduct of this study, you may contact the Protocol Officer for Ethics in Research, University of Ottawa, Tabaret Hall, Room 154, 550 Cumberland Avenue, Ottawa, ON K1N 6N5, 613-562-5387, or ethics@uottawa.ca.

~~~~~ Consent ~~~~~ I give the researchers permission to use video segments in educational or research presentations.

**YES** \_\_\_\_\_ **NO** \_\_\_\_\_

My signature below confirms that I **DO** consent to my infant’s participation in this study, conducted by Dr. Christopher Fennell of the School of Psychology. I also understand that our participation in this study is entirely voluntary and that we may refuse to participate or withdraw from the study at any time. I have received a copy of this consent form for my records.

Name of Infant (please print): \_\_\_\_\_  
 Name of Parent(s) (please print): \_\_\_\_\_

Parent's signature: \_\_\_\_\_ Date: \_\_\_\_\_  
Parent's signature: \_\_\_\_\_ Date: \_\_\_\_\_  
Researcher's signature: \_\_\_\_\_ Date: \_\_\_\_\_

**Appendix B: Background Questionnaire**

Subject Code: \_\_\_\_\_ Date: \_\_\_\_\_

**Family Background Questionnaire**

*\*Note that all language questions are part of our Language Background Questionnaire*

Your child's sex/gender: M / F / Other

Your child's date of birth: \_\_\_\_\_

Your child's due date: \_\_\_\_\_

Has your child had one or more ear infections since they were born?

Does your child currently have an ear infection? Yes / No

Yes / No

Does your child have a known hearing or vision problem? Yes / No

If so, please explain:

\_\_\_\_\_

Do you have any concerns about your child's development? Yes / No

If so, please explain:

\_\_\_\_\_

**Siblings**

Sibling 1 Date of birth: \_\_\_\_\_ Sibling 2 Date of birth: \_\_\_\_\_ Sibling  
3 Date of birth: \_\_\_\_\_ Sibling 4 Date of birth: \_\_\_\_\_

**Parent(s)**

Parent 1: Sex/gender: M / F / Other

Sex/Gender: Sex/Gender: Sex/Gender: Sex/Gender:

M / F / Other M / F / Other M / F / Other M / F / Other

How many years did you spend in formal schooling (e.g., elementary school, junior high school, high school, college, university)?

\_\_\_\_\_

What is the highest degree you earned?

High school diploma or equivalency (GED)  College or trades degree or certificate  
 Bachelor's degree  
 Master's degree

Doctorate  
 Professional (MD, JD, DDS, etc.)  Other specify  
 None of the above (less than high school)

#### FONCTIONS EXÉCUTIVES ET BILINGUISME 47

If applicable, partner's sex/gender: M / F / Other

If applicable, how many years did your partner spend in formal schooling (e.g., elementary school, junior high school, high school, college, university)?

\_\_\_\_\_

If applicable, what is the highest degree your partner earned?

High school diploma or equivalency (GED)  College or trades degree or certificate  
 Bachelor's degree  
 Master's degree

Doctorate  
 Professional (MD, JD, DDS, etc.)  Other specify  
 None of the above (less than high school)

Which of these categories best describes your total combined family income for the past 12 months? This should include income (before taxes) from all sources.

Less than \$5,000  \$5,000 through \$11,999  \$12,000 through \$15,999  
 \$16,000 through \$24,999  \$25,000 through \$34,999  \$35,000 through \$49,999  
 \$50,000 through \$74,999  \$75,000 through \$99,999  \$100,000 and greater  
 Don't know  
 No response

What is your current postal code? \_\_\_\_\_





|  |  |  |  |  |  |
|--|--|--|--|--|--|
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |