2023

The Phonetic Specificity of British English-Learning Infants' Word Form Recognition in Their First Year of Life

Ratnage, Paul

https://pearl.plymouth.ac.uk/handle/10026.1/21794

http://dx.doi.org/10.24382/5121
University of Plymouth

UNIVERSITY OF PLYMOUTH

The Phonetic Specificity of British English-Learning Infants' Word Form Recognition in Their First Year of Life

by

PAUL RATNAGE

A thesis submitted to the University of Plymouth

in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

May 2023

Copyright statement

Acknowledgments

Firstly, thank you to all the parents and infants who volunteered their time to take part in research in the BabyLab and made the data collection for this thesis possible. I am extremely grateful to my Director of Studies, Caroline Floccia, for introducing me to the wonderful world of infant research as an undergraduate, and for all her guidance, mentorship, and understanding during the PhD. Thanks also to Thierry Nazzi for all his feedback and Lionel Granjon for providing the stimuli used in Experiment 1. I am also appreciative to the University of Plymouth for funding the PhD program.

I am very thankful to everyone who has worked in the BabyLab and Tech Office throughout my thesis, especially the many amazing placement students for all their tireless efforts in recruitment, testing, and the day-to-day running of the lab. I would like to acknowledge the help of Sophie Olver for help with recording the audiovisual stimuli and Martyn Atkins for programming the eye-tracking experiments.

Special thanks go to Samantha Durrant, for encouraging me to apply for the PhD, and her invaluable advice throughout. Also, to Katja Stärk for all her helpful encouragement. Thanks also to Karen Green and Jodie Cook. I also wish to thank my parents for all their support, even if they still don’t really understand what I do!

Finally, this thesis would not have been possible without the endless love and support from my incredible wife, Sarah. Thank you for everything you have done to help get me through this PhD, especially when I really struggled during the pandemic. Your belief in me has kept me going. We can go on holiday now!

Author's declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee. Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

This study was financed with the aid of a studentship from the University of Plymouth.

Publications

Ratnage, P., Nazzi, T., & Floccia, C. (in press). Vowels and consonants matter equally to British English-learning 11-month-olds' familiar word form recognition. *Journal of Child Language*, 1-24. https://doi.org/10.1017/s0305000923000223

In preparation

Ratnage, P. & Floccia, C. (in prep) Audiovisual speech benefits British English-learning 12-month-olds' word recognition.

Presentations at conferences

Ratnage, P, Nazzi, T., Granjon, L., & Floccia, C. (2018). The impact of phonetic and intensity changes on word recognition in British English-learning 5-month-olds. *Poster presentation at the 3rd Summer School on Infant Studies on Language Development in Europe (ISOLDE), Potsdam, Germany.*

Ratnage, P, Nazzi, T., Granjon, L., & Floccia, C. (2018). The impact of phonetic and intensity changes on word recognition in British English-learning 5-month-olds. *Poster presented at the 21st International Congress of Infant Studies* (*ICIS), Philadelphia, USA.*

Ratnage, P, Nazzi, T., Granjon, L., & Floccia, C. (2018). The impact of phonetic and intensity changes on word recognition in British English-learning 5-month-olds. *Spoken presentation at the 3rd Lancaster Conference on Infants and Early Child Development* (*LCICD), Lancaster, UK.*

Ratnage, P, Nazzi, T., & Floccia, C. (2019). Do British English-learning 11-month-olds identify consonant and vowel changes in familiar words? *Poster presentation at the 15th Experimental Methods in Language Acquisition Research* (*EMLAR), Utrecht, Netherlands.*

Ratnage, P, Nazzi, T., & Floccia, C. (2019). Do British English-learning 11-month-olds consonant and vowel changes in familiar words? *Poster presentation at the 4th Workshop on Infant Language Development* (*WILD), Potsdam, Germany.*

Ratnage, P, Nazzi, T., Christodoulou, F., Kemp, A., Hitchin, L., & Floccia, C. (2019). Do British English-learning 11-month-olds identify consonant and vowel changes in familiar words? *Spoken presentation at the Child Language Symposium* (*CLS), Sheffield, UK.*

Word count of main body of thesis: 56377

Signed:

Date: 04/05/23

The Phonetic Specificity of British English-Learning Infants' Word Form Recognition in Their First Year of Life

Paul Ratnage

Abstract

Consonants and vowels have been proposed to have distinct functions in speech perception: a consonant bias for lexical processing and a vowel bias for syntactic/prosodic processing (Nespor et al., 2003). Research in adults has consistently demonstrated that consonants have a privileged role in various lexical-level experiments across most languages. However, cross-linguistic differences have been found in the developmental trajectory of the consonant bias. For example, whilst French-learning infants display a consonant advantage in lexical processing tasks by their first birthday (e.g., Poltrock & Nazzi, 2015), British English-learning infants show an equal sensitivity to consonants and vowels until the age of 30 months (e.g., Floccia et al., 2014). Although the lexical and/or the acoustic-phonetic properties of an infant's native language have been hypothesised to explain such variations, additional cross-linguistic tests of the consonant bias and its potential links to these factors are required. The present thesis explored this by using two experimental paradigms to further examine the phonetic specificity of British English-learning infants' word form recognition at the onset of lexical acquisition. Experiments 1 to 3 established an equal preference for consonant and vowel mispronunciations of familiar word forms, presented either in isolation or in list form, in 5-, 11-, and 12-month-old infants using the head-turn preference procedure. Experiments 4 and 5 used an eye-tracking methodology to measure whether the congruent presentation of audio and visual speech signals led to a consonant bias in 12-month-olds' word form recognition. An audiovisual benefit was found, with infants discriminating between phonetic mispronunciations, but only when they viewed a speaker articulate alterations of a single familiar word form. Additionally, neither acoustic factors (Experiment 1) nor lexical factors (Experiments 2 to 5) were found to influence infants' preferences. Together, the results of this thesis provide further evidence that initial lexical processes vary cross linguistically.

List of contents

List of tables

List of figures

1 Introduction

A considerable body of infancy research has found that exposure to native language input shapes perception. Although infants are born with the ability to discriminate many of the speech sounds found in the world's languages (for a review, see Aslin et al., 1998), their perception narrows to focus on native vowel contrasts by 6 months (Kuhl et al., 1992; Polka & Werker, 1994) and native consonant contrasts by 10 to 12 months (Best et al., 1988; Werker & Lalonde, 1988; Werker & Tees, 1984). Alongside acquiring the sound structure of their native language, infants are beginning to recognise frequently heard common and proper nouns. Parental reports suggest that 6- to 9-months-olds understand the meaning of many words (Frank et al., 2017). This is supported by experiments which show that infants not only prefer listening to familiar words over unfamiliar words during their first year of life (Hallé & de Boysson-Bardies, 1994; Hallé & de Boysson-Bardies, 1996; Swingley, 2005; Vihman et al., 2004), but can also identify a referent upon hearing a familiar word (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999).

To understand how infants build phonological representations of their first words, research has typically examined their sensitivity to alterations or removals of a word's consonants rather than its vowels (Hallé & de Boysson-Bardies, 1996; Swingley, 2005; Vihman et al., 2004). However, consonants and vowels have been proposed to have contrasting roles in language processing (Nespor et al., 2003). Specifically, consonants are said to provide more information relating to the lexicon, whereas vowels would provide more information concerning syntactic and prosodic processing. Adult studies in many languages (including English, French, Spanish, Italian, and Dutch) have almost always demonstrated this greater dependence on consonants over vowels in both lexical learning and lexical access (with the possible exception of tone languages, for a review, see Nazzi & Cutler, 2019). Therefore, given that the privileged role of consonants in adult lexical processing appears to be found irrespective of the language's phonological inventory, such a bias could assist infants in learning their native language. Specifically, a consonant bias could help facilitate both word learning and word recognition in infants.

Despite considerable support for the presence of the consonant bias in adults, its developmental trajectory remains unclear. Originally proposed as being an innate bias (Nespor et al., 2003), research has consistently found cross-linguistic differences in the weight infants attribute to consonant and vowel information in lexical processing tasks (Nazzi et al., 2016). For example, French-learning infants display a consonant advantage in word form recognition by 11 months (Poltrock & Nazzi, 2015). In contrast, word form recognition in infants learning British English, the language at the focus of this thesis, appears to be equally impacted by consonants and vowels until beyond 24 months (Mani & Plunkett, 2007). These divergent findings have been hypothesised to be related to the lexical (Keidel et al., 2007) and/or acoustic-phonetic (Floccia et al., 2014) properties of an infant's native language. However, the existing evidence does not allow for a conclusion that the consonant bias develops due to either of these factors. To gain more strength for either hypothesis, additional cross-linguistic tests of the consonant bias and its potential links to lexical and/or acoustic-phonetic factors are required.

This thesis aims to further examine the development of the consonant bias. Using two different research methodologies, it will examine whether British English-learning infants rely on consonants or vowels in familiar word form recognition during their first year of life. Firstly, Chapter One consists of a literature review, focussing initially on infants' phonological acquisition and word form recognition abilities, before providing an overview of evidence of the consonant bias in both adults and infants. Chapter Two presents a series of head-turn preference studies, which will investigate how alterations to the consonants and vowels of familiar word forms influence their identification in 5- (Experiment 1), 11- (Experiment 2), and 12-month-olds (Experiment 3). It will also evaluate how acoustic or lexical factors may impact such recognition.

Chapters Three and Four will then consider whether providing audiovisual cues may alter infants' abilities to detect consonant and vowel changes to familiar words. Chapter Three is a further literature review of infants' abilities to audiovisually match phonemes, how incongruent phonemes change infants' attention towards a speaker's face, as well as how audiovisual cues may influence phonetic discrimination and word recognition. Chapter Four will then present two eye-tracking studies in 12-month-olds, examining how audiovisual consonant and vowel mispronunciations modify their face scanning behaviour and propensity for displaying a consonant bias when presented with lists of words (Experiment 4) or repetitions of a single word (Experiment 5). Finally, the General Conclusion will evaluate the findings presented in this thesis, assess how they link to each developmental hypothesis, and detail ideas for further research into understanding the acquisition of the consonant bias.

2 Chapter One

2.1 Infants' early sensitivity to speech

Human neonates are born prepared for language acquisition. Regions of the frontal and temporal cortex, critical to language processing in adults, are already activated when listening to spoken language in the first few days and weeks of life (e.g., Altvater-Mackensen & Grossmann, 2016; Dehaene-Lambertz et al., 2002; May et al., 2018; Peña et al., 2003; Shultz et al., 2014). Similar to lateralisation in adults, the neural responses to language in infants are frequently found to be more pronounced in amplitude in the left hemisphere (Altvater-Mackensen & Grossmann, 2016; Dehaene-Lambertz et al., 2002; Peña et al., 2003; Shultz et al., 2014), although such lateralisation has not always been replicated in newborns (May et al., 2011; May et al., 2018; Perani et al., 2011; Taga et al., 2007) Furthermore, the auditory system begins to develop during the third trimester of pregnancy, with the capacity to hear appearing between the 24th to 25th gestational week (Birnholz & Benacerraf, 1983). This affords the developing foetus an exposure to a range of environmental sounds, such as a low-pass filtered variation of their mother's voice (Lecanuet & Schaal, 2002). Although detailed phonological information is absent in the womb, neonates are exposed to prosodic features, allowing them to extract information relating to the intonation and rhythm of their native language, even before they are born.

Beginning at birth, newborns prefer listening to human speech over other categories of sound. For example, Vouloumanos and Werker (2007) used a high-amplitude sucking procedure to measure neonates' preference for speech compared to non-speech analogues that matched both temporal and spectral parameters of spoken language. In this method, an infant is presented with an auditory stimulus which is contingent upon their sucking on a pacifier. Once the infant's sucking decreases to a predefined threshold, a new auditory stimulus is played. An increase, rather than no change, in the infant's sucking rate is interpreted as evidence for a preference for the auditory change. They results showed that newborns' sucking significantly increased when listening to speech rather than a non-speech analogue. Such a preference for language has been found in comparison to a range of non-speech sounds, including tones, filtered speech, scrambled speech, and monkey calls (Shultz & Vouloumanos, 2010; Vouloumanos & Werker, 2007; Vouloumanos et al., 2010). A recent meta-analysis by Issard et al. (2021), analysing 38 experiments on infants' preferential listening to speech over non-speech sounds during the first year of life, reported that human language is reliably preferred over other categories of sound, including non-vocal and other vocal sounds, as well as artificial and natural sounds. Furthermore, this language preference is maintained throughout an infant's first year.

The effects of prenatal experience of language heard in the womb are behaviourally observed in infants' speech perception. Newborns differentiate between their mother's voice and another female's voice (Beauchemin et al., 2011; DeCasper & Fifer, 1980), and prefer familiar over novel passages or melodies read by their mother during late pregnancy (DeCasper & Spence, 1986; Partanen et al., 2013). Neonates also selectively listen to the language they are exposed to when in the womb. To illustrate, American English newborns can discriminate between speech in their native language and speech in an unfamiliar language, such as French (Mehler et al., 1988; Moon et al., 1993). Similarly, neonates exposed to two languages during gestation show a comparable preference for each language (Byers-Heinlein et al., 2010). Newborns even appear to cry in the prosodic pattern of the language heard in utero, such as German newborns crying in a falling melody and French infants crying in a rising melody (Mampe et al., 2009).

Neurologically, hearing their native language elicits increased cortical activity in newborns in comparison to an unfamiliar language (Fava et al., 2014; Sato et al., 2012). Such brain activation is more pronounced when listening to forward speech over backward speech if spoken in a newborn's native language but not when spoken in a non-native language (May et al., 2018). There is even evidence of distinct hemispheric specialisations in newborns for native and unfamiliar languages. Specifically, a left hemisphere lateralisation has been found in newborns listening to their native language, whereas right hemisphere lateralisation is associated with hearing a non-native language (Sato et al., 2012; Vannasing et al., 2016). This suggests that native language is processed by distinct neural networks from birth.

Newborns are also able to detect variations in the segmental rhythm between distinct languages. Rhythmic classes of languages are classified in relation to the timing units that describe their rhythms and are traditionally categorised as syllable-timed (e.g., French and Italian), stress-timed (e.g., Dutch and English), and mora-timed (e.g., Japanese). Newborns can differentiate between stress-timed and syllable-timed languages (Mehler et al., 1988; Moon et al., 1993), and between stress-timed and mora-timed languages (Nazzi et al., 1998; Ramus et al., 2000). In contrast, when presented with two languages that belong to the same rhythmic class (e.g., stress-timed English and Dutch), neonates are unable to discriminate between languages (Nazzi et al., 1998).

Newborns can differentiate between sets of sentences from stress-timed languages (e.g., English and Dutch) and sets of sentences from syllable-timed languages (e.g., Italian and Spanish) but are unable to do so when the sentences in each set are from different rhythmic classes (Nazzi et al., 1998). Thus, newborns are sensitive to both their native language and the rhythmic differences that exist between languages.

2.2 Phonological acquisition in the first year of life

Cross-language perception research has established that an adult's ability to categorise and differentiate speech sounds is associated with the phonological characteristics of their native language (see Section 2.4 for a detailed discussion on consonants and vowels). Indeed, whilst some non-native contrasts are easier to detect than others (Best et al., 1988), adults often struggle to perceive differences between phonemes that do not contrast phonologically in their native language (e.g., Best & Strange, 1992; Faris et al., 2018; Flege & Eefting, 1987; Guion et al., 2000; Tyler et al., 2014). For example, native speakers of Japanese struggle to discriminate between the non-native English /r/ and /l/ consonantal contrast in comparison to native speakers of English (e.g., Guion et al., 2000). These limitations can be difficult to overcome even with increased exposure (e.g., Pallier et al., 1997). From a developmental perspective, infants must therefore become attuned to the phonetic categories that are associated with their native language. Specifically, infants must develop a sensitivity to which speech sounds communicate differences in meaning (e.g., recognising the contrast between *b* and *t* in order to recognise *bin* from *tin*).

2.2.1 Initial sensitivity to phonetic categories

Extensive research over the last 50 years has demonstrated that very young infants can differentiate between both native and non-native phonemic contrasts. In relation to consonant perception, Eimas et al. (1971) used a high-amplitude sucking technique to present 1- and 4-month-old infants with pairs of synthesised stimuli that differed along the Voice Onset Time (VOT) continuum. As a measure of the acoustic differences between stop consonants of contrasting voicing categories (e.g., /p, b, t/), VOT refers to the interval (in ms) between the release of an oral constriction and when the vocal folds begin to vibrate (Lisker & Abramson, 1964). For instance, the phoneme boundary between /b/ and /p/ has a VOT of around 25ms (Wood, 1976), such that adult listeners categorically perceive everything below the interval as /b/ and everything above as /p/ (Liberman et al., 1957). Eimas et al. (1971) found that both 1- and 4-month-olds discriminated changes that crossed the English phoneme boundary (20ms to 40ms; /ba/ to /pa/) but did not discriminate changes that occurred within a phoneme boundary (e.g., 60ms to 80ms; /pa/ to /pa/). This finding importantly resembled adult like categorical perception of consonants (Liberman et al., 1957). Furthermore, very young infants can not only discriminate voicing contrasts, but also the manner of articulation (e.g., Eimas & Miller, 1980), and place of articulation (e.g., Eimas, 1974).

Subsequent research has demonstrated that infants can differentiate between non-native consonant contrasts. For example, English-learning 2-month-olds can differentiate between the non-native Czech contrast /ʒ/–/r̝/ (Trehub, 1976), and Kikuyu-learning infants can differentiate between the non-native English-like /b/–/p/ contrast (Streeter, 1976). English-learning 4- and 6-month-olds can discriminate between both the

Tamil /n̪ /–/ɳ/ and /l/–/ɭ/ and Filipino /n/–/ŋ/ contrasts (Sundara et al., 2018; but see Narayan et al., 2010, for a failure to find discrimination of the Filipino contrast). At 6 months, English-learning infants can detect the difference between two different Hindi consonantal contrasts, namely /ʈa/–/ta/ and /th/–/dh/ (Werker et al., 1981), and Japanese-learning 6-month infants can discriminate the English /r/ and /l/ consonants (Kuhl et al., 2006). Thus, despite little or no exposure, very young infants appear to be able to discriminate between consonants not found in their native language.

Brain imaging studies have also revealed perceptual discrimination of native and non-native consonants (e.g., Dehaene-Lambertz & Dehaene, 1994; Mahmoudzadeh et al., 2013; Peña et al., 2012; Rivera-Gaxiola et al., 2005). For instance, a mismatch negativity (MMN) refers to an automatic identification of perceptual change that occurs without conscious awareness or a behavioural response (Näätänen et al., 2007). Peña et al. (2012), using electroencephalography (EEG), found that 9-month-old Spanish-learning infants displayed a MMN to hearing both a native (/b/-/d̥/) and non-native (/d̥/-/ɖ̥/) consonantal change. Similarly, results using event-related potentials (ERPs) have found that 3-month-old infants detect the phonemic change in digitised consonants /ba/ and /ga/ in less than 400ms (Dehaene-Lambertz & Dehaene, 1994). Therefore, there is both behavioural and neurophysiological evidence that young infants can perceive consonantal contrasts that do not belong to their native language inventory.

As with consonants, an initial ability to discriminate between vowels has been demonstrated in young infants. Trehub (1973), again using a high-amplitude sucking procedure, presented 1- to 4-month-old infants with the vowels /a/, /i/, and /u/, phonemic contrasts that are found in most language systems (Maddieson, 1984). The vowels were either spoken in isolation (e.g., /a/ - /i/) or in a consonant-vowel context (e.g., /ta/ - /ti/). The infants were successful in detecting the vowel changes in both scenarios, suggesting that an initial sensitivity to vowels is present from an early age. Additional research has replicated this finding, establishing that infants can discriminate /a/, /i/, and /u/ from birth (Kuhl & Miller, 1982; Kujala et al., 2004; Sebastián-Gallés & Bosch, 2009). Furthermore, Swoboda et al. (1976) tested 2-month-old infants' discrimination of the more subtle vowel contrast of /i/-/ɪ/, in both between-categories (i.e., /i/-/ɪ/, /ɪ/-/i/), and within-categories (i.e., /i/-/i/, /ɪ/-/ɪ/). Infants perceived the vowel changes continuously, discriminating both between- and within-category vowel contrasts equally well. This result is comparable to the continuous perception of vowels found in adults (Fry et al., 1962).

Young infants can also differentiate between vowels that are not found in their native language. For instance, 2-month-old English-learning infants can successfully differentiate between the non-native French vowels /a/–/ã/ (Trehub, 1976). At 4 months, the non-native Catalan /e/–/ɛ/ vowels are discriminated by Spanish-learning infants (Bosch & Sebastián-Gallés, 2003). German-learning 6- to 8-month-olds can discriminate the English /ɛ/–/æ/ vowel contrast, while the same aged English-learning infants can discriminate the German /u/–/y/ vowel contrast (Polka & Bohn, 1996). These behavioural findings are further supported by findings from neurophysiological research. For example, ERPs reveal that 1- to 5-day-old Finnish-learning newborns show a MMN response to the vowels /i/-/y/ (Cheour-Luhtanen et al., 1995). Similarly, 6-month-old Finnish-learning infants display a MMN when discriminating between a native Finnish /e/-/ö/ vowel contrast, as well as a non-native Estonian /e/-/õ/ vowel contrast (Cheour et al., 1998).

Thus, as with consonantal contrasts, infants appear to be able to discriminate between vowels, irrespective of whether they are found in their native language.

These findings show that infants initially possess impressive phoneme discrimination abilities. Whilst not all phonemes have been found to be successfully discriminated, including consonants and vowels present in their native language (e.g., Finnish newborns and their native /i/–/ɨ/ vowel contrast, Cheour-Luhtanen et al., 1995; 3-week- to 6-month-old Swedish-learning infants and their native /ɑ/–/ɔ/, Kuhl et al., 1992; 6- to 8-month-old American English-learning infants and their native /ɹ/–/l/ distinction, Kuhl et al., 2006), infants are perceptually sensitive to a wide range of native and non-native phonetic contrasts from birth.

2.2.2 Perceptual attunement to phonetic categories

Across the first year of life, an infant's initial ability to discriminate many of the world's speech contrasts shifts towards those found in their native language. This process has been referred to as perceptual attunement (or perceptual narrowing). To illustrate, Werker and Tees (1984) assessed whether English-speaking adults and English-learning infants could discriminate between the non-native Hindi /Ta/-/ta/ and Salish /k'i/-/q'i/ contrasts. Infants were tested using a conditioned head-turn paradigm. This task trains infants to adjust their heads in the direction of an audio speaker and a reinforcer (e.g., an animatronic toy) when they hear a change in an auditory stimulus. A head-turn in response to a change from the Hindi /Ta/ to /ta/ would be interpreted as infants differentiating between phonetic contrasts. It was found that, at 8- to 10-months-old, English-learning infants could successfully differentiate between both the Hindi and Salish contrasts. However, 10- to 12-month-old English-learning infants were unable to discriminate between either non-native contrast, a pattern that was additionally demonstrated in the adults tested. In contrast, no decline in discrimination was found for native contrasts in the same aged Hindi- and Salish-learning infants (Werker et al., 1981; Werker & Tees, 1983, 1984). This finding suggests that, between the ages of 10 and 12 months, the initial sensitivity for non-native consonants is significantly reduced, with infants instead showing adult-like perceptual limitations.

The pattern of perceptual attunement demonstrated by Werker and Tees (1984) has been replicated in infants for a wide range of consonantal contrasts in both behavioural and neurophysiological tasks (e.g., Best et al., 1995; Peña et al., 2012; Rivera-Gaxiola et al., 2005; Tsushima et al., 1994). For example, Japanese-learning infants' discrimination of the non-native English /r/ and /l/ declines between 6 to 8 months and 10 to 12 months. In comparison, no decline is shown for their native /w/ and /y/ consonantal contrast (Tsushima et al., 1994). Similarly, the MMN response in Spanish-learning infants to the non-native consonantal change /d̥/-/ɖ̥/ at 6 months is no longer exhibited at 12 months, whereas a MMN is found for a native /b̥/-/d̥/ consonantal change at both ages (Peña et al., 2012). The pattern of perceptual narrowing has even been shown in infants' discrimination of sign language contrasts (Palmer et al., 2012). Furthermore, although some differences may occur during the first year of life, research suggests that even bilinguals have learned the consonant contrasts found in each of their native languages by the age of 12 months (e.g., Burns et al., 2007).

Attunement to vowels follows a similar, if not earlier, developmental pattern to consonants, such that infants' perception reflects their native vowel categories by approximately 6 months (Tsuji & Cristia, 2014). This narrowing of speech perception has been researched in the context of the perceptual magnet effect (Kuhl, 1991; Kuhl, 1993; Kuhl et al., 1992), in which discrimination accuracy is found to be worse for better vowel examples than for poorer vowel examples. For instance, when participants have been familiarised with a prototypical /i/ vowel, they are unsuccessful in discriminating non-prototypical instances of that vowel. In contrast, if participants are first familiarised with a non-prototypical /i/ vowel, they can successfully discriminate a prototypical version. Thus, a prototypical vowel attracts, or acts as a magnet, to non-prototypical instances, consequently leading them to be not as distinguishable. This effect has also been found in both 6-month English-learning infants and English-speaking adults (Grieser & Kuhl, 1989; Kuhl, 1991).

Kuhl et al. (1992) used the perceptual magnet effect to examine how linguistic experience may impact the perception of vowels. In a conditioned head-turn task, English- and Swedish-learning 6-month-olds' discrimination of native and non-native vowels was examined. Infants were presented with synthesised prototypes of the English /i/ and the Swedish /y/, along with modified, acoustically comparable, non-prototypical variations of the vowels. They found that infants demonstrated a perceptual magnet effect but, crucially, only for their native vowel categories. Specifically, English-learning infants failed to differentiate between variants of their native /i/ that were acoustically similar to the prototype compared to variants that were less similar to the prototype. However, Swedish-learning infants were able to discriminate both prototypical and non-prototypical versions. The opposite pattern was found for the Swedish /y/, with only Swedish-learning infants' discrimination of non-prototypical variants impacted by similarity to the prototypical version. This result demonstrated that linguistic experience altered infants' phonetic perception, with vowel categories appearing to be attuned to those in their native language by 6 months.

Although under researched in comparison to consonantal contrasts, several studies have supported Kuhl et al.'s (1992) finding of a decline in the ability to differentiate between non-native vowels in comparison to native vowels (e.g., Bosch & Sebastián-Gallés, 2003; Jansson-Verkasalo et al., 2010; Polka & Bohn, 2011; Polka & Werker, 1994). For example, English-learning infants can differentiate between two non-native German vowel contrasts (/U/–/Y/ and /u/–/y/) at 4 months but show a decline in discrimination at 6 to 8 months, and a failure to discriminate at 10 to 12 months (Polka & Werker, 1994). Catalan and Spanish monolinguals, as well as Spanish-Catalan bilinguals, can all discriminate the Catalan /e/ and /ε/ vowels at 4-months-old. In comparison, only Catalan monolinguals can successfully differentiate the contrast at 8 months (Bosch & Sebastián-Gallés, 2003). Similarly, Finnish-learning infants' MMN, as measured by EEG, declines for non-native vowels between the age of 6 and 12 months (Jansson-Verkasalo et al., 2010). Indeed, a meta-analysis by Tsuji and Cristia (2014), compiling the results of 18 experiments examining the perceptual attunement to vowels, concluded that the perception of vowels has typically narrowed to those found in an infants' native language by the age of 6 to 9 months.

Reorganisation of native phoneme perception is therefore seen at a younger age for vowels (Kuhl et al., 1992) than consonants (Werker & Tees, 1984). One explanation for this difference in age of acquisition may be due to an increased exposure to vowels. Although typologically most languages contain more consonants than vowels (Ladefoged

& Disner, 2012), approximately 45% of the speech stream in stress-timed languages, 50% of the speech stream in syllable-timed languages, and 55% in mora-timed languages is comprised of vowels (Ramus et al., 1999). Thus, by 6 months, infants have been exposed to more vowel tokens than consonant tokens, providing them with more opportunities to analyse their native vowels compared to their native consonants (Maye et al., 2002). Vowels are also higher in intensity and longer in duration in comparison to consonants (Repp, 1984). Consequently, both foetuses (Granier-Deferre et al., 2011) and neonates (Benavides-Varela et al., 2012; Bertoncini et al., 1988) perceive vowels more clearly than consonants. This again may provide infants with more exposure to their native vowels. There is also evidence that newborns' vowel perception is impacted by the language exposed to in the womb, such as English-learning and Swedish-learning infants showing perceptual attunement to their native English /i/ or native Swedish /y/ vowels respectively (Moon et al., 2013). Finally, vowels carry the pitch characteristics of infant directed speech, so they may be initially more attention grabbing to infants (Fernald & Kuhl, 1987).

It is important to note that there are some native consonant contrasts, such as /d/-/ð/ in French-learning infants (Polka et al., 2001; Sundara et al., 2006), which remain difficult for infants to discriminate. Similarly, some non-native vowel contrasts continue to be differentiated by older infants, such as the German /u/-/y/ and English /ɛ/–/æ/ contrasts in English-learning and German-learning 10- to 12-month-olds respectively (Polka & Bohn, 1996). There are a few exceptions where no decline in discrimination of non-native consonants has been shown, including 12- to 24-month-old English-learning infants and English-speaking adults showing no perceptual decline for Zulu clicks (Best et al., 1988). One explanation for the ability to perceive non-native consonants is provided by Best's

(1995) Perceptual Assimilation Model, according to which discrimination performance depends upon how perceptually similar non-native phonemes are to native phonemes. If there is no clear similarity to a native consonant or vowel, then the non-native phoneme may still be perceived.

There is evidence that, alongside the decline in the ability to discriminate non-native categories, speech perception of native categories significantly improves as infants age. Kuhl et al. (2006) found that, while both 6- to 8-month-old and 10- to 12-month-old English-learning infants could differentiate between a native /r/ and /l/ contrast, discrimination performance improved significantly in 10- to 12-month-olds. A longitudinal experiment by Rivera-Gaxiola et al. (2005) found a more robust MMN response to native consonantal contrasts between infants tested at 7 months and 11 months. Comparable increases in the MMN response have been reported for native vowel contrasts when examined at 6 and 11 months (Cheour et al., 1998). This suggests that, as infants develop, a stronger discriminatory response is found for contrasts that are lexically relevant in their native language. Furthermore, individual differences in discrimination of native contrasts can predict later language acquisition (for a review, see Cristia et al., 2014). Kuhl et al. (2008) found that an increased MMN to native phonetic contrasts at 7.5 months, compared to an increased sensitivity to non-native phonetic contrasts, predicted a larger vocabulary size at 18 and 24 months. Therefore, a more advanced native phonetic expertise may predict an accelerated attainment of a native language.

2.3 Word form recognition in the first year of life

Before infants achieve the milestone of producing their first word, usually in the second year of life, they must first understand some of the words of their native language. To achieve this, infants must segment words from the speech stream, remember such words, and then form word-referent links. Typically, the prerequisite for *knowing* a word involves a knowledge of its sound form (i.e., the sequence of its phonemes), as well as an awareness of its syntactic properties and semantic reference. However, in relation to young infants, *word form recognition* is often used to highlight that, even if they do not yet understand the meanings of words, infants may know their sound forms either in isolation or segmented from fluent speech (Swingley, 2009).

2.3.1 Segmenting words from speech

Before infants can begin to acquire the words found in their native language, they must first learn to extract them from the continuous speech. Speech segmentation should constitute a difficult problem for infants. Speech is both rapid and continuous, missing conclusive pauses to clearly identify word boundaries (Klatt, 1979), with adult listeners relying on several probabilistic cues to successfully segment speech (Cutler, 2012). Words also rarely occur in isolation. Infant speech corpora have revealed that the majority of speech directed towards 6- to 12-month-old infants is composed of multi-word utterances, with estimates indicating that less than 10% of speech that is either directed towards or heard by infants consists of isolated words (Brent & Siskind, 2001; Morgan, 1996; van de Weijer, 1998). Although single-word utterances may have a significant contribution to language development (Brent & Siskind, 2001; Lew-Williams et al., 2011), infants must nevertheless begin to segment words from the speech stream to learn their native language.

One behavioural method used to measure infants' word segmentation abilities is the head-turn preference procedure (Fernald, 1985), a paradigm which explores infants' abilities to differentiate between pairs of auditory stimuli. In a typical head-turn preference experiment, infants are seated on a parent's lap in a three-sided booth. Red lights and loudspeakers, which are controlled by an experimenter viewing a live video feed, are located to the left and right of the infant, along with a centrally located green light. At the start of each sequence, the central green light will flash until the infant's head is directed towards it, at which point it will extinguish and one of the red side lights will begin to flash. Once the infant has turned their head towards the flashing side light, an auditory stimulus is played through its associated loudspeaker. This auditory stimulus will play either until completion or until the infant turns their head away from the light for a set period, typically two seconds. The looking time towards the light is used as a measure of an infant's interest in the auditory stimulus presented to them. Increased looking times toward one set of sounds over another is interpreted as infants discriminating between auditory stimuli (Nelson et al., 1995).

In a series of seminal studies, Jusczyk and Aslin (1995) used the head-turn preference procedure to examine 6- and 7.5-month-old American English-learning infants' abilities to detect the sound patterns of words in fluent speech. Infants were first familiarised with isolated repetitions of two target words (e.g., "cup" and "dog"). Looking times were then recorded for passages consisting of sentences that either did or did not include the familiarised words in utterance-initial, -medial, and -final positions. The results showed that 7.5-month-olds, but not 6-month-olds, listened longer to passages containing the words they had previously been exposed to. This suggests that the older infants recognised the familiarised words when they were embedded in fluent speech. A further study presented 7.5-month-old infants with passages that included two target words before testing their preference for listening to the isolated target words and two novel words (e.g., “bike” and “feet”). Again, infants were found to attend more to the target words they had previously been familiarised with compared to novel words. Therefore, by 7.5 months, infants were able to extract segments of auditory information from fluent speech, treated as words by adult listeners, despite such segments having no attached meaning.

Furthermore, findings from ERP studies show infants’ segmentation abilities. ERP research has an advantage over the head-turn preference paradigm due to not requiring any preference or behavioural change from the infant. Instead, an infant’s EEG is recorded whilst they passively listen to speech and attend to an unrelated silent video to maintain their attention. In relation to word segmentation, comparing the time course of ERPs between a familiarised and unfamiliarised word can provide a measure of when and if an infant’s brain recognises words. For example, Kooijman et al. (2005) first familiarised Dutch-learning 10-month-olds with isolated tokens of words unlikely to be familiar to them. ERPs were then recorded while the infants attended to a series of sentences containing either the familiarised word or an unfamiliarised word. The results revealed that ERPs to familiar words were more negative in amplitude in comparison to novel words, with a divergence occurring approximately between 350-500ms following word onset. This suggests that infants not only recognised the previously familiarised words but were also displaying a recognition response before the offset of the word. This result has been found in several other languages, including in French- (Goyet, et al., 2010) and German-learning (Männel & Friederici, 2013) 12-month-olds. Therefore, both behavioural and electrophysiological studies demonstrate the early ability to segment speech.

Further research has focussed on the cues that infants use to segment words from speech. The findings from such studies reveal that, from 6 months onwards, infants display an increased capacity to identify words in continuous speech, utilising different cues to help them achieve speech segmentation. For instance, 7.5-month-old American English-learning infants can use the strong/weak pattern of stress found in English to segment words from fluent speech (Jusczyk, Houston, et al., 1999). At 8 months, infants can use the transitional probabilities between syllables - a conditional probability statistic that measures the likelihood of event Y given event X in the sequence XY - to segment artificial speech into word forms (Saffran, Aslin, et al., 1996). From 8 to 9 months, infants begin to use language specific cues to segment words. These include the syllabic-based unit in syllabic-based languages (e.g., French; Goyet et al., 2010) or the strong/weak stress patterns found in some stressed-based languages (e.g., English; Jusczyk, Houston, et al., 1999). Infants also begin to show increased sensitivity to native segmental cues such as coarticulation (Johnson & Jusczyk, 2001), allophones (Mattys & Jusczyk, 2001), and phonotactics (Gonzalez-Gomez & Nazzi, 2013). The overall conclusion from such studies is that infants can extract word forms from fluent speech during their first 12 months and possess multiple cues to assist them with the task.

2.3.2 Early word form recognition

Evidence of word form recognition in the first year of life has been provided by experiments evaluating the listening times for isolated familiar word forms compared to unfamiliar words or pseudowords. The rationale behind such research is that, if an infant differentiates between familiar and unfamiliar words, it demonstrates a recognition of the sound patterns of frequently heard words. In particular, studies on infants aged less than 6-months-old have focussed on the recognition of their own name, which is likely to be a salient word form in their environment. For example, an infant's name is frequently used in isolation to attract their awareness (Morikawa et al., 1988) and is typically placed in the most salient utterance position in sentences (Durkin et al., 1982; Newport et al., 1977). Similarly, a prosodic pause is often found when an infant's name is spoken in a sentence (e.g., "*Ollie, don't do that!*") and will often draw their attention (Durkin et al., 1982).

Mandel et al. (1995) were the first to demonstrate that infants can recognise the isolated sound pattern of their own name when they are just 4.5-months-old. Using a modified version of the head-turn preference procedure, American English-learning infants were presented with isolated repetitions of their own name and isolated repetitions of other names. Specifically, infants heard their own name (e.g., *Sarah*), a name with a corresponding stress pattern to the infant's own name (e.g., *Michael*), and two names consisting of the opposite stress pattern to the infant's own name (e.g., *Kathleen* and *Nicole*). Such stress-matched and stress-mismatched foils were used to explore if infants only had partial representations of their own names. For instance, differentiation between their own name and a name with a different stress pattern would only be possible if infants have acquired the global prosodic features of their own name. Mandel et al. (1995) found that the infants attended longer when listening to repetitions of their own name, irrespective of whether the foil had the same stress pattern or a different stress pattern. Similar findings have been found in French- (Bouchon et al., 2015) and British English-learning (Delle Luche et al., 2017) 5-month-olds. This suggests that, from the age of 4.5 months, infants have already begun to recognise the frequently occurring sound pattern of their own name. Furthermore, such recognition is detailed enough that it is identified in contrast to names with prosodically similar sound patterns.

Infants' abilities to recognise their own name has been supported by additional neurological research. Parise et al. (2010) reported a higher ERP amplitude in German-learning 5-month-olds when they listened to their own name compared to an unfamiliar name which differed in the first phoneme. This differentiation was found to occur within 100-380ms of the stimulus onset. This indicated that infants could recognise their own name over an unfamiliar name from the first phoneme alone when the first phoneme of the unfamiliar name differed. Additionally, this experiment examined if own name recognition guided infants' attention. Having heard their own name or an unfamiliar name, infants were show pictures of novel objects. The amplitude of the ERP was found to be higher when infants were presented with objects that had been preceded by their own name compared to objects preceded by an unfamiliar name. This suggests that infants not only recognise if their name is being spoken but can use this familiar word form as a social cue to focus their attention.

Early word form recognition may even provide a salient cue for segmenting words from fluent speech. Bortfeld et al. (2005) used the head-turn preference procedure to first familiarise 6-month-old infants with familiar-name target passages (sentences in which their own name was followed by a novel word) and alternate-name target passages (sentences containing a novel name followed by a different novel word). For example, an infant named Maggie would hear sentences such as “The bell on *Maggie’s bike* was really loud” (familiar-name target trials) and “She put *Hannah’s cup* back on the table” (alternate-name target trials). In a test phase, looking times to the novel words were measured to examine if infants showed better recognition for familiar-name target over alternate-name target words. The results demonstrated that, even though they had been equally exposed to both words, infants only recognised the word when it was preceded by their own name and not the word when it was preceded by an unfamiliar name. This finding has been replicated in 8-month-old French-learning infants, with the word *maman* (meaning *mummy*) assisting infants with segmenting novel words from an artificial language (Mersad & Nazzi, 2012). These findings demonstrate that young infants recognise some repeatedly heard names and use such familiar words as initial anchors during speech segmentation.

Research in slightly older infants has explored whether they recognise other frequently heard word forms beyond their own name. For instance, Hallé and de Boysson-Bardies (1994) used the head-turn preference procedure to examine 11- and 12-month-old French-learning infants’ word form recognition. Infants listened to two separate lists of bisyllabic words from their native language that were either familiar or rare. The familiar words, such as *bonjour* (meaning *hello*) and *lapin* (meaning *bunny*), were words which were likely to be heard frequently in their surroundings. The rare words, such as *busard* (meaning *harrier*) and *caduc* (meaning *obsolete*), were closely matched in their phonetic structure but were likely to be infrequently spoken, if at all, in an infants’ environment.

Infants were found to listen longer to the familiar words over rare words, despite the absence of any visual referents or training of the words heard. This suggests that infants display word form recognition for frequently heard words at this age. Similar findings have been found using EEG at 11 months (Thierry et al., 2003). Such word form recognition has been replicated in several languages, including in British English- (Vihman et al., 2004), Dutch- (Swingley, 2005), and Italian- (Vihman & Majorano, 2017) learning 11-month-olds. However, this effect does not appear in British English-learning 9- (Vihman et al., 2004) or 10-month-olds (Vihman et al., 2007), indicating that stable representations of familiar word forms are not developed until infants reach 11 months. Thus, by the time they reach their first birthday, infants demonstrate recognition of some familiar words, even if they may not yet understand their meaning.

2.3.3 Early word comprehension

Several experiments have investigated whether infants are able to associate word forms to a visual referent during the first year of life. The most frequently used procedure for measuring infants' word comprehension is the intermodal preferential looking (IPL) paradigm. This method presents infants a series of trials of side-by-side images or videos, each depicting different objects or events, on a screen. Each trial lasts several seconds, midway through which the infant hears a target word, presented either in isolation or in a carrier sentence, that matches only one of the images or videos on the screen. Infants' eye movements during the trial are recorded, either by video camera or by automatic eye trackers. Infants are deemed to have learnt the association between an object and a label if they look longer towards the target in comparison to the distractor after hearing its auditory label.

Using the IPL task, Tincoff and Jusczyk (1999) showed 6-month-old infants side-by-side silent videos of their mother and father while auditory labels of the words *mummy* or *daddy* were played. Infants were found to orientate longer towards the videos of their mother when hearing *mummy* and towards their father when hearing *daddy*. This suggests that, even at this young age, infants were able to correctly link sound patterns with meaning. A second experiment found that infants looked equally towards videos of unfamiliar females and males when hearing the same labels. This demonstrates that infants did not extend the word *mummy* to include all females, or the word *daddy* to include all males. Using the same method, Tincoff and Jusczyk (2012) found that 6-month-old infants looked longer at either an image of an adult's hand or foot when the associated label was played. Thus, infants' word comprehension appears to extend beyond the specific one-to-one associations of *mummy* and *daddy* to more complex categories of objects. Furthermore, given the infants had not been previously exposed to the images of the hand or foot presented, the finding indicates the infants were also able to generalise and extend their direct experience (i.e., hearing their own hands and feet being labelled by their parents) to a perceptually comparable object.

Relatedly, Bergelson and Swingley (2012) used the IPL procedure to explore whether 6- to 9-month-old infants have an understanding for common nouns. Infants were shown pictures of familiar body parts (e.g., mouth) and foods (e.g., banana), either presented as two discreet images (paired-picture trials), or as an image depicting several items (scene trials). Paired-picture trials displayed items from across the two categories (i.e., one body part and one food), whereas the scene trials depicted only within-category items. A target image was named in each trial, spoken in the familiar voice of the infant's parent, with eye-tracking used to determine if infants gazed towards the specified picture. The results revealed that both 6- and 9-month-old infants were not only adept at recognising the correct item in paired trials but were capable of identifying a named item in the more complex scene trials. Thus, infants appear able to link a word form with a visual referent during the second half of their first year.

Additional research has supported the finding that infants show word comprehension from 6 months onwards. Bergelson and Swingley (2015) found that 6- to 16-month-old infants' recognition of words was above chance when attending to a named object in side-by-side videos. They additionally demonstrated that infants' word recognition abilities follow a nonlinear trajectory, remaining stable from 6 months onwards until a boost in performance is found at 13 to 14 months. Furthermore, infant noun comprehension at 9 months, as measured with IPL, is correlated with the words that their parents report them understanding (Syrnyk & Meints, 2017). Using EEG, Parise and Csibra (2012) found that 9-month-olds display a N400 mismatch effect, an ERP component linked to lexical-semantic processing, when a displayed image of an object is incongruous to a label just spoken by their mother. This provides neurological evidence of early word comprehension. Beyond nouns, Nomikou et al. (2019) reported that 10-month-old infants displayed an increase in target gaze having heard an action verb. For example, infants looked longer at the image of a banana over building blocks following the sentence "What can you eat?". Finally, word learning studies suggest that infants can be trained to associate a novel label with a novel object during their first year (e.g., Friedrich & Friederici, 2017; Pruden et al., 2006; Twomey & Westermann, 2018). Together, these findings provide additional support to infants' early word recognition abilities, demonstrating that infants are also showing the beginnings of word comprehension before their first birthday.

2.3.4 Phonetic specificity of early word form recognition

Research has explored whether infants can detect incorrectly pronounced versions of familiar words. The ability to identify such mispronunciations would provide evidence that infants' early word form recognition is phonetically specified. In American English-learning infants, there is evidence that phonetic information is encoded for familiar words from the age of 7.5 months. In the previously discussed segmentation studies by Jusczyk and Aslin (1995), infants were also familiarised with mispronunciations of monosyllabic words which had their onset consonant altered by a single phonetic feature. For example, *cup* was mispronounced as *tup*. Infants then heard the correctly produced words in a series of sentences, along with sentences containing a novel target word. If infants' word representations are phonetically detailed at this early age, then, during this test phase, it would be expected that they would attend equally to sentences containing either the correctly pronounced familiarised word or an unfamiliarised word. Jusczyk and Aslin (1995) found that infants displayed a comparable preference for both series of sentences. This suggests that, at 7.5 months, American English-learning infants' words are phonetically specified such that even a single phonetic feature change will impact their later recognition.

Hallé and de Boysson-Bardies (1996), building on their previous finding that infants preferred listening to familiar words over unfamiliar words (Hallé & de Boysson-Bardies, 1994), investigated the phonetic specification of infants' lexical representations in a series of head-turn preference studies. French-learning 11-month-old infants were again found to prefer listening to familiar bisyllabic words compared to unfamiliar words. This effect was found when the familiar words were mispronounced by either changing the place of articulation or voicing of the initial consonant (e.g., *bonjour* pronounced as *vonjour* or *ponjour*). Similarly, a non-significant tendency to prefer listening to familiar words was shown when the second syllable of the onset consonant was altered (e.g., *bonjour* pronounced as *bongour*). Only when completely removing the initial consonant of the familiar word (e.g., *bonjour* pronounced as *oonjour*) did infants show no preference for mispronounced familiar words over unfamiliar words. In a further experiment, infants did not show any preference between the correctly pronounced familiar words (e.g., *bonjour*) and mispronunciations on either the initial consonant (e.g., *vonjour*) or the second syllable of the onset consonant (e.g., *bongour)*. Together, these findings indicate that French-learning 11-month-old infants' early lexical representations may not be specified enough to recognise both correct pronunciations and certain mispronunciations of familiar words.

The incongruity of the finding that detailed phonetic representations for words are found in American English-learning 7.5-month-olds (Jusczyk & Aslin, 1995) but not in French-learning 11-month-olds (Hallé & de Boysson-Bardies, 1996) may be explained by several factors. For example, Hallé and de Boysson-Bardies (1996) suggested that a change may occur in infants' representations of familiar word forms in the second half of their first year of life. In particular, early word form representations may initially be phonetically specified but, as infants' lexicons grow, their representations may become less specified and more global by 12 months.

Language differences may also be a factor, specifically the dissimilar lexical stress systems found in English and French. Lexical stress refers to the degree of respiratory energy placed on a syllable (Ladefoged & Disner, 2012). This can occur in different syllable positions within a word and is associated with vowel reduction in unstressed syllables. In English, the stress pattern is typically trochaic, whereby the stress is placed on a word's first syllable. Indeed, the majority of disyllabic words in English have a stress that is word initial (Cutler & Carter, 1987). In comparison, French has an iambic stress pattern, with the (non-lexical) stress placed on the second syllable of a word. This distinction may impact the phonetic specification of the onset consonant mispronunciations used in Jusczyk and Aslin (1995) and Hallé and de Boysson-Bardies (1996). Thus, French-learning infants may not attend to the onset consonant due to the iambic nature of their native language. In contrast, English-learning infants may attend more to the initial consonant of a word, given that it is stressed. This is supported by the finding that English-learning 9-month-olds attend longer to lists of trochaic English words (e.g., *provide*) in comparison to lists of iambic (e.g., *idle*) English words (Jusczyk et al., 1993). Similarly, English-learning 7.5-month-olds segment trochaic bisyllabic words (e.g., *kingdom*) but not iambic bisyllabic words (e.g., *guitar*; Jusczyk, Houston, et al., 1999). Therefore, word stress may explain why English-learning infants' words appeared phonetically specified whereas French-learning infants' words did not.

Vihman et al. (2004) also used the head-turn preference procedure to measure English-learning 11-month-olds' preference for lists of familiar bisyllabic words over unfamiliar words. Again, the words were pronounced either correctly (e.g., *bubbles*), with an alteration at the onset of the word (e.g., *mubbles*), or with an alteration at the onset of the second syllable (e.g., *bummles*). They first found that infants attended longer to familiar words in comparison to unfamiliar words. In contrast, infants showed no preference when listening to onset alterations over unfamiliar words, suggesting that changing the manner of articulation of an onset consonant impacted word recognition. Interestingly, infants preferred listing to alterations at the onset of the second syllable over unfamiliar words, indicating weak phonological encoding. However, this finding was only found on trials presented in the second half of the experiment, with infants showing no preference during the first half. Vihman et al. (2004) proposed that infants at first failed to recognise the mispronunciations of the familiar words but, following exposure, a preference for the mispronunciations over unfamiliar words emerged. Therefore, it appears that English-learning 11-month-olds' word recognition is impaired by mispronunciations in the onset consonants of both syllables found in familiar bisyllabic words, although it may take infants longer to recognise them.

Following their finding that English-learning infants detected mispronunciations at the onset of the second syllable, but only during the second half of the experiment, Vihman et al. (2004) reanalysed the data from Hallé and de Boysson-Bardies (1996). This reanalysis revealed that 11-month-old French-learning infants' preference for familiar words with mispronunciations on the onset consonant in comparison to unfamiliar words emerged only in the second half of the experiment. Therefore, French-learning infants appeared to detect mispronunciations at the onset consonant, whereas English-learning infants detected mispronunciations for medial consonants. This result may also be clarified in relation to the stress systems used in English and French (Vihman et al., 2004). Similarly, it may explain why French-learning infants' recognition is blocked when mispronunciations occur at the onset of the second syllable but not at the onset of a word (Hallé & de Boysson-Bardies, 1996). Thus, the phonetic recognition of words in 11-month-olds may be impacted by the manner of articulation and their native language.

Further evidence that early word form recognition contains phonetic detail has been provided by Swingley (2005). In a series of experiments, Dutch-learning 11-month-olds' preference for listening to familiar monosyllabic CVC word forms (e.g., /vut/, meaning *foot*) over pseudowords was found to disappear when the familiar words were altered on either the onset consonant (e.g., /but/) or the offset consonant (e.g., /vuk/). This demonstrates that infants appeared to have a detailed enough phonetic representation of the onset and offset of the words that such alterations appeared to block their word form recognition when compared with pseudowords. However, additional experiments revealed that infants listened significantly longer to correct pronunciations over onset mispronunciations but attended equally to correct pronunciations and offset mispronunciations. According to Swingley (2005), this later finding suggests that, although infants are sensitive to offset alterations, they can still recognise the phonological similarity between correctly pronounced words and offset mispronunciations. Such results highlight that, at 11 months, Dutch-learning infants appear to encode the phonological details of the onset of familiar words, but that the encoding of the offset of familiar words is weaker.

The evidence presented so far demonstrates that, during their first year of life, infants can not only extract words forms from fluent speech (Jusczyk & Aslin, 1995) and recognise word forms (Mandel et al., 1995), but are also beginning to display word comprehension (Bergelson & Swingley, 2012). Furthermore, by 12 months, infant word recognition is impacted by mispronunciations, particularly when they occur on the stressed syllable of their native language (Hallé & de Boysson-Bardies, 1996; Vihman et al., 2004). Thus, infants' early word form recognition appears to be phonetically specified, to some extent. The following section will explore this further by examining the differing role that consonants and vowels may have in the phonetic specificity of infants' first words.

2.4 Consonants and vowels

As the two basic phonological categories present in each of the world's languages (Ladefoged & Disner, 2012), consonants and vowels are integral to the structure of speech. There are an estimated 600 unique consonants and 200 different vowels found in all the world's approximate 7,000 languages. Each language consists of a unique subset of these phonemes (Ladefoged & Disner, 2012). As previously discussed, infants acquire their native phonological categories at different stages of development. Native vowel categories are acquired by approximately 6 months (Kuhl et al., 1992), with native consonant categories learned later at around 10 to 12 months (Werker & Tees, 1984). However, beyond differences in their age of acquisition, consonants and vowels can be distinguished on several levels.

In terms of production, consonants are speech sounds that are created with either a complete or partial obstruction of the vocal tract. Table 1 outlines the consonants found in British English.

Table 1. *Consonants in British English (adapted from Roach, 2004).*

	Bilabial		Labio-dental		Dental		Alveolar		Post-alveolar		Palatal		Velar		Glottal	
Unvoiced (-V) Voiced (+V)	*-V*	*+V*	*-V*	*+V*	*-V*	*+V*	*-V*	*+V*	*-V*	*+V*	*-V*	*+V*	*-V*	*+V*	*-V*	*+V*
Plosive	p	b					t	d					k	g		
Affricate									tʃ	dʒ						
Nasal		m						n						ŋ		
Fricative			f	v	θ	ð	s	z	ʃ	ʒ					h	
Approximant		(w)								r		j		w		
Lateral Approximant								l								

Consonants can be classified in terms of their voicing, based on whether the vocal cords vibrate during production. *Voiced* consonants (e.g., /b, d, m/) are sounds when the vocal cords vibrate, whereas *voiceless* consonants (e.g., /k, p, s/) are sounds made when the vocal cord does not vibrate (Yavaş, 2011). Consonants can vary in relation to their place of articulation, which describes where in the vocal tract an obstruction in the airstream occurs. The predominant articulators that cause these obstructions are the lips and the tongue (Ladefoged & Johnson, 2015). Examples of places of articulation in English include *bilabial* (/b, m, p/), made when the lips come together, *palatal* (/j/), formed when the front of the tongue is raised toward the hard palate behind the alveolar ridge, and *alveolar* (/t, d, s, z, n, l/), sounds produced when the tongue is positioned behind the upper front teeth (Yavaş, 2011). Consonants can also differ in their manner of articulation, which is the magnitude and type of obstruction occurring in the vocal tract. The articulators may either shut the oral tract for a succinct or extended period, considerably reduce the space in the oral tract, or they may approach each other to alter the shape of the oral tract (Ladefoged & Johnson, 2015). Examples of manners of articulation in English are *stops* (e.g., /p, b, t/), which are produced when the vocal tract is completely blocked resulting in no airflow, *fricatives* (e.g., /f, v, s/), when a partial airflow is squeezed through a narrow gap between the articulators, *affricates* (/tʃ/, /dʒ/), which involve an oral closure and then a slow release, and *nasals* (/m, n, ŋ/), sounds produced when the air in the vocal tract is released through the nose rather than the mouth. Both fricatives and nasals are examples of continuants as, unlike stops, they allow airflow to continue during production (Yavaş, 2011). Additionally, stops and continuants can be differentiated by their energy, with stops having a large and quick increase in amplitude at release compared to continuants (Stevens & Blumstein, 1981).

In contrast to consonants, vowels are produced without any obstruction to the air passage (Ladefoged & Disner, 2012). Figure 1 outlines the vowels found in British English.

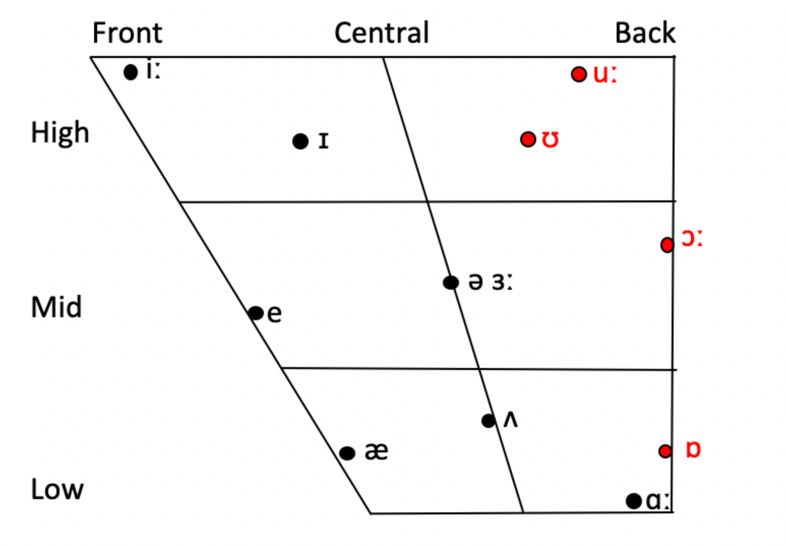


Figure 1: *Vowels in British English. Vowels in red are rounded (adapted from Roach, 2004).*

Given that the articulation of vowels involves no contact between the articulators, both place and manner of articulation do not apply in the production of vowel sounds. Similarly, vowels are typically voiced, so are not categorised in terms of being voiced or voiceless. Instead, vowels can be distinguished by the position of the lips, and both the height and front-back position of the tongue (Yavaş, 2011). In relation to the lips, some vowels are produced when the lips are either *rounded* (e.g., /uː/) or *unrounded* (e.g., /ɪ/). The height of the body of the tongue is defined as being the tongue's position relative to the roof of the mouth, being either *high* (e.g., /iː/), *medium* (e.g., /e/) or *low* (e.g., /æ/) based upon the vowel being articulated. The part of the tongue involved in producing vowels can also be classified in relation to its backness, being either *front* (e.g., /æ/), occurring when the tongue is relatively close to the roof of the mouth, *back* (e.g., /ɔː/), when the tongue is near to the upper area of the vocal tract, and *central* (e.g., /ʌ/), with the location falling in between the front and back positions (Ladefoged & Johnson, 2015). Vowels can additionally be classed as monophthongs or diphthongs. Monophthongs are steady state vowels, in which there is no noticeable alterations in phonetic quality. Diphthongs are complex vowel sounds which form a single syllable, and include a change from one vowel to another, such as the /<u>eɪ</u>/ in *tape* (/t<u>eɪ</u>p/; Yavaş, 2011).

The differences in production when articulating consonants and vowels means that the two phonemes are acoustically distinguishable. Speech sounds consist of the acoustic properties of time, amplitude, and frequency. Time relates to the duration of a specific speech sound. For instance, the length of the alveolar stop /t/ is shorter than that of the alveolar fricative /s/. Amplitude is the degree of subglottal air pressure, with greater variation in air pressure resulting in a louder sound. Finally, frequency, sometimes described as pitch, is the speed at which part of a sound wave recurs. Measured in hertz (Hz), frequency is made up of fundamental frequency (f0), expressing the amount of vocal cord vibrations when making speech sounds, and formant frequency, relating to the configuration of the vocal tract (Yavaş, 2011). Time, amplitude, and frequency have been found to differ between consonants and vowels. Vowels are typically longer in duration and are higher in intensity in comparison to consonants (Repp, 1984). Consequently, vowels are more salient than consonants. The increased clarity of vowels means they are less likely to be misperceived than consonants (Bond & Garnes, 1980). In relation to frequency, consonants are categorised as rapid changes in frequency over time, whereas vowels have more steady-state frequencies (Cole & Scott, 1974; Liberman et al., 1967; Stevens & Blumstein, 1978).

Perceptual differences also exist between consonants and vowels, demonstrated by findings from phonetic categorisation tasks. This methodology presents participants with an artificial continuum, which begins with an unambiguous phoneme, passes through an ambiguous section, before reaching a final unambiguous phoneme (e.g., between /b/ and /p/). The listener's task is to either discriminate between the phonemes on such a continuum or to decide whether a sound token from the continuum matches one phoneme endpoint better than the other (McQueen, 1996). Consonants are typically found to be perceived more categorically than vowels in such phonetic categorisation tasks. For example, listeners can successfully differentiate between contrasting stop consonant categories but are unable to do so when they are from the same category. Similarly, consonant categorisation decisions are determined by the closest endpoint phoneme on the continuum (Liberman et al., 1957), a result that has been found in infants by the end of their first month of life (Eimas et al., 1971). In comparison, vowels are not as categorically defined. Instead, a more continuous representation is found along a continuum for vowel identification, as well as an increased sensitivity within categories (Fry et al., 1962; Pisoni, 1973). Relatedly, there is evidence that consonants are processed faster than vowels in both visual-word recognition (Vergara-Martínez et al., 2011) and auditory (Delle Luche et al., 2014) domains.

Neuropsychological research from a range of methods provides evidence that consonants and vowels are distinctly represented in the brain, with different neural responses triggered during processing. For example, divergent patterns of errors when using consonants and vowels have been found in patients with brain damage. Caramazza et al. (2000) reported on two patients suffering with conduction aphasia, a condition characterised by an impairment in the capability to repeat phrases or words following brain injury. One patient exhibited errors in reproducing consonant segments of words, whereas the other showed the opposite pattern, struggling with vocalic segments. Similarly, Ferreres et al. (2003) described a patient with alexia, a reading disorder following damage to the brain, whose reading of nonwords was more impacted by vowels than consonants. Functional magnetic resonance imaging (fMRI) studies have revealed distinct brain activation patterns for consonants and vowels. Carreiras and Price (2008) found different neural responses to consonants and vowels in both reading and lexical decision tasks. For instance, asking participants to change the vowels in spoken words (e.g., prim**a**v**e**ra to prim**e**v**a**ra) caused greater activation in a right middle temporal area, a brain region linked to prosodic processing. In comparison, changing consonants in spoken words (e.g., pri**m**a**v**era to pri**v**a**m**era) caused greater activation in a right middle frontal area, a brain region linked with response inhibition. The results from ERPs have also found distinctive electrophysiological responses to consonants and vowels in word processing (Carreiras et al., 2007), in lexical decision tasks (Carreiras et al., 2008), and in abstract rule learning (Monte-Ordoño & Toro, 2017). Therefore, consonants and vowels appear to have different neural representations and activate discrete neural responses.

Finally, the number of consonants and vowels not only varies from language to language, but consonants typically outnumber vowels cross-linguistically. To illustrate, English has a total of 36 phonemes, consisting of 24 consonants and 12 vowels (plus 13 diphthongs); French has 39 phonemes, which are made up of 20(+2) consonants and 17 vowels (plus 4 diphthongs); Italian has 31 phonemes, comprised of 24 consonants and 7 vowels. Indeed, most languages usually contain over 20 consonants and at least 5 vowels (Maddieson, 1984). There are even a few rare examples of systems, such as Hawaiian (8 consonants, 5 vowels) and Rotokas (6 consonants, 5 vowels), where relatively few consonants will still outnumber vowels (Nespor et al., 2003). Similarly, there are instances of languages, such as Tashlhiyt Berber (Ridouane, 2008), which have lexicons that contain many words comprised entirely of consonants. Systems that contain more vowels than consonants are less frequent. Such languages include Danish, with 15 consonants and 21 short and long vowels, and Swedish, which contains 16 consonants and 17 vowels.

2.5 Division of labour hypothesis

In addition to the dissimilarities between consonants and vowels previously outlined, Nespor et al. (2003) have suggested that a ‘division of labour’ is also found between the two phonetic categories. Specifically, they theorised that the asymmetry between consonants and vowels leads to individual's possessing two distinct processing biases in relation to speech perception. Firstly, due to consonants being more beneficial than vowels in recognising and learning the words of a language, there is a consonant bias for processing lexical information. Secondly, due to vowels being more helpful than consonants in identifying the grammatical and rhythmic properties of language, there is a vowel bias for processing syntactic and prosodic information.

Nespor et al. (2003) identified several reasons why consonants and vowels have specific processing biases. In relation to a consonant bias, consonants typically outnumber vowels in most of the world's languages, with only a few exceptions of systems where vowels are more numerous than consonants (e.g., Danish). This numerical difference means that consonants may provide more distinctions in relation to articulatory features compared to vowels. The distinctiveness of consonants is further supplemented by the fact that they tend to disharmonise in many languages. Specifically, if the consonants within a word share the same feature, one will typically alter its value (Mehler et al., 2006). In comparison, vowels tend to harmonise in most languages, thus reducing their distinctiveness.

Consonants, rather than vowels, can also comprise morphological roots in certain languages, again making them more distinct. Semitic languages, where the morphological roots of words are composed solely of consonants, provide a good example of the distinctiveness of consonants. For instance, the root *k-t-b*, with the meaning *to write* in Arabic, generates contrasting words or word forms based upon the vowels that separate the consonants (Nespor et al., 2003). The combination of vowels placed between the consonants allow the formation of different words that are connected by their meaning (e.g., *kitab*, meaning *book*; *katib*, meaning *writer*, etc.). In contrast, vowels signify morphological patterns (McCarthy, 1985). To illustrate, the combination of vowels placed between the root k-t-b, allow the formation of different words that are connected by their meaning (e.g., kitab, meaning book; katib, meaning writer, etc.). However, there are no languages that have lexical roots formed by vowels alone. This suggests that sequences of consonants, rather than vowels, are more beneficial for recognising words.

In comparison, the vowel bias is based on the observation that vowels are more important than consonants in marking prosodic and syntactic regularities. Vowels typically differ more than consonants in their frequency, duration, and intensity (e.g., Liberman et al., 1967; Repp, 1984). Consequently, vowels carry most of the prosodic detail of a language, including its rhythm, intonation, and phrase boundaries. To illustrate, the syllabic repertoire of a language can be predicted by the number of vowels that are contained in the speech stream (Ramus et al., 1999). This syllabic complexity is associated with the length of the most frequent words found in a language (Mehler & Nespor, 2003). Furthermore, prosody is also linked to the syntactic properties of speech, such as word order, constituency, and disambiguation (Nespor & Vogel, 1986). Importantly, infants are not only sensitive to prosody from birth (e.g., Christophe et al., 2001; Nazzi et al., 1998), but can use such detail to infer the syntactic structures and rules of their native language (Christophe et al., 1997; Morgan & Demuth, 1996). For instance, infants can use prosodic cues to segment words from fluent speech (Jusczyk, Houston, et al., 1999), discover the word order of function and content words in their native language (Gervain & Werker, 2013), and learn syntactic constituency (Hawthorne & Gerken, 2014). Such findings indicate that the information carried by vowels may assist with syntactic acquisition.

Nespor et al.’s (2003) hypothesis that there are distinct functional biases for consonants and vowels has been supported by a range of experimental data in both adults and children. The scope of the present thesis is concerned with the developmental acquisition of the consonant bias and how this may impact lexical processing in infants. Therefore, the following section will primarily focus on how consonants may impact word recognition and word learning in adults before evaluating evidence of a consonant bias in infants.

2.6 Evidence of the consonant bias in lexical processing in adults

Support for the existence of consonant bias in lexical processing has been provided by studies in adults that have demonstrated a greater dependence on consonants over vowels in tasks associated with both lexical access and lexical learning (Nazzi & Cutler, 2019). These include word reconstruction tasks (e.g., van Ooijen, 1996), lexical decision tasks (e.g., New et al., 2008), the segmentation of artificial languages (e.g., Bonatti et al., 2005), and word learning (e.g., Creel et al., 2006). Furthermore, these findings have been demonstrated across several different languages.

2.6.1 Word reconstruction tasks

Word reconstruction tasks, developed by van Ooijen (1996), ask adult participants to listen to pseudowords that have been created by modifying real words by a solitary segment. Each pseudoword participants hear can be converted into a real word by altering either a single consonant or a single vowel. For example, the pseudoword *kebra* could be transformed into the real word *zebra* if an alteration were made to the initial consonant, or into the real word *cobra* if an alteration were made to the initial vowel. The participant's task is to quickly change the pseudoword they have heard into a real word by converting a single phoneme. This task can be implemented by allowing participants a free choice to the phoneme they alter or restricting them to changing either the consonant or vowel. In English, van Ooijen (1996) observed that participants were more likely to make vowel changes to the pseudowords when given a free choice. Similarly, participants were both slower and made more errors (e.g., replacing the wrong type of phoneme) in word reconstructions when they were asked to make a consonant change in comparison to being asked to alter a vowel. Furthermore, the effect was irrespective of the number of consonants and vowels contained in each pseudoword. This finding has been replicated using fMRI (Sharp et al., 2005), with left anterior frontal activation higher when performing consonant word reconstructions. Therefore, in word reconstruction tasks, consonants are more useful than vowels.

The consonant bias found in the word reconstruction task in English-speaking participants (van Ooijen, 1996) has been replicated in other languages, including different variations of Spanish (Cutler et al., 2000; Marks et al., 2002), Japanese (Cutler & Otake, 2002), and Dutch (Cutler et al., 2000). Beyond demonstrating that the consonant bias is not language specific, the findings from these studies provide further insights into the role of consonants in lexical processing. Firstly, the consonant bias does not appear to be impacted by the number of phonemes contained in each language. To illustrate, Cutler et al. (2000) found that both Spanish-speaking (a language with a higher consonant-vowel ratio and high distinctiveness between its vowels) and Dutch-speaking (a language with a similar number of vowels and consonants and many similar sounding vowels) participants typically altered vowels when given a free choice and were both slower and less accurate when asked to change consonants rather than vowels. Secondly, the cross-language findings suggest that phonology does not impact the consonant bias, with comparable results found in English (van Ooijen, 1996), a language that has both vowel reductions and lexical stress, Spanish (Cutler et al., 2000; Marks et al., 2002), which contains no vowel reduction, and Japanese (Cutler & Otake, 2002), which has no lexical stress. Finally, the findings suggest that dialect experience does not impact the consonant bias, with Spanish dialects typically fluctuating more in consonants than vowels in comparison to English dialects (Nazzi & Cutler, 2019). Therefore, word reconstruction tasks support the assertion that a consonant bias in lexical processing is language universal.

2.6.2 Lexical decision tasks

Findings from priming studies have demonstrated the advantage of consonants over vowels in lexical decision tasks. In a masked priming lexical decision experiment, participants must decide whether an uppercase letter string is a real word or a nonword as quickly as possible. This target item is preceded, usually by around 30 to 50ms, by a lowercase prime that is embedded either side of a forward mask (usually #####) and the target, such that participants are unaware of the prime. Responses to a target (e.g., *TABLE*) are quicker when the masked prime is a related word (e.g., the identical word *table*) than an unrelated word (e.g., *zebra*), a finding known as a priming effect (Forster & Davis, 1984).

New et al. (2008) used the masked priming lexical decision task to present French-speaking adults with primes that preserved the consonant or vocalic information of the target word. Four types of primes were presented 50ms before the target: consonant-related (e.g., *duvo-DIVA*), vowel-related (e.g., *rifa-DIVA*), identity-related (e.g., *diva-DIVA*), and unrelated (e.g., *rufo-DIVA*). Although identity-related primes produced the quickest response, participants were faster at correctly identifying real words that had been preceded by a consonant-related prime compared to vowel-related primes. Indeed, responses to vowel-related primes were comparable to responses to unrelated primes. Furthermore, this pattern of results has been replicated in the auditory modality in both French- and English-speaking adults (Delle Luche et al., 2014). These findings substantiate the advantage that consonantal information has over vocalic information in aiding visual word recognition.

An additional study by New and Nazzi (2014), using the same primes and words as New et al. (2008), manipulated the duration of the presentation of primes to further examine the consonant bias in French-speaking adults. Firstly, they tested the orthographic and phonological/lexical nature of the consonant bias in word recognition. Previous research suggests that only orthographic effects are found for primes displayed at 33ms and only phonological effects are found for primes displayed at 60ms (Ferrand & Grainger, 1993). New and Nazzi (2014) demonstrated that neither consonant-related nor vowel-related primes produced a priming effect when presented at 33ms. However, when primes were presented at 66ms, consonant-related primes produced faster reaction times than vowel-related primes. This suggests that a consonant priming effect is occurring at a phonological/lexical level rather than an orthographic one.

Priming differences between consonants and vowels have also been demonstrated in studies in which participants are presented with a prime and target in which two letters have been transposed. For instance, Perea and Lupker (2004) found a priming effect in Spanish participants when two consonants were transposed (e.g., *caniso*-*CASINO*) but not when two vowels are transposed (e.g., *anamil*-*ANIMAL*). Similarly, Duñabeitia and Carreiras (2011) explored the role of consonants and vowels on the relative position priming effect. This effect is based on the finding that target word recognition is quicker when the priming word retains some of the target's letters in their relative position (Grainger et al., 2006). They found relative position priming in Spanish-speaking adults when primes were composed only of consonants (e.g., *csn*-*CASINO*) but not when primes contained only vowels (e.g., *aio*-*CASINO*).

2.6.3 Word form segmentation

Consonants have been found to have an advantage in artificial language segmentation tasks. For example, Bonatti et al. (2005) familiarised French-speaking participants with artificial speech streams formed of CVCVCV pseudowords (e.g., *biduka*). The transitional probabilities between the syllables of the pseudowords were manipulated such that they were carried by either the consonant or vowel. Both adults (Saffran, Newport, et al., 1996) and 8-month-old infants (Saffran, Aslin, et al., 1996) can use transitional probabilities between syllables to segment artificial speech into word forms. In one condition, Bonatti et al. (2005) used a transitional probability of 1.0 between the consonants found in each word, with the vowels varying. For instance, the consonants *b_d_k_* would always appear in the same sequence, with the blanks completed by different variations of vowels. In another condition, the transitional probability between vowels was 1.0, with the consonants instead varying (e.g., *_o_ε_y*). In a test phase, participants were then asked to identify which one of two items in a pair belonged to the language they had heard. It was observed that participants could successfully identify words contained in the artificial language when the transitional probability between the consonants of such words was 1.0. However, they were unable to perform this task when the transitional probability between the vowels of the words in the artificial language was 1.0. This finding suggests that adults can only successfully track transitional probabilities when they are carried by consonant sequences but not when they are carried by vowel sequences. This again provides further support for a consonant bias in lexical tasks, revealing that consonants are more beneficial when segmenting words from speech.

Relatedly, a series of experiments by Toro et al. (2008) in Italian-speaking adults investigated the functional differences between vowels and consonants in the ability to segment words and identify syntactic rules. They reported that adults again only extracted words from an artificial speech stream when their transitional probabilities were carried by consonants and not vowels. This finding extends the beneficial role of consonants in word segmentation found in Bonatti et al. (2005) to another language. Moreover, Toro et al. (2008) observed that when vowels in the artificial language followed an ABA syntactic rule, in which the first and last vowels of a CVCVCV word were identical (e.g., *badeka*), participants were able to generalise this rule to identify new words belonging to the language. However, when the consonants in the artificial language followed the same

ABA syntactic rule (e.g., *binebo*), participants showed no evidence of rule-learning, performing at chance levels when identifying new words. Together, such research reinforces Nespor et al.'s (2003) claim that consonants and vowels have dissimilar functions in speech processing, demonstrating that consonants are advantageous in segmenting speech and vowels are more beneficial at a syntactic level.

2.6.4 Word learning tasks

Consonants have also been found to be more advantageous than vowels when learning new words. For instance, Creel et al. (2006) taught American adults the labels for a series of novel shapes before testing their ability to select the correctly labelled shape from four recently learned alternatives. They found that participants were more likely to be confused by the recently learned words when they shared consonants (e.g., *suba* and *sabo*) than when they shared vowels (e.g., *nasi* and *tagi*). This showed that participants relied more on consonants for lexical identification in comparison to vowels. Furthermore, this pattern of results was independent of the number of consonants and vowels contained in each word, although a weaker consonant bias was found when vowels were in an onset position.

Similarly, Havy et al. (2014) taught French-speaking adults labels for pairs of novel objects that either differed by one consonant (e.g., */pyv/* and */tyv/*) or one vowel (e.g., */pos/* and */poes/*). In an eye-tracking test phase, participants were shown the image pairings, with one being named. No significant differences in overall learning were found, with participants performing equally well in looking at the correct object in consonant and vowel contrasting pairs. However, participants were quicker at detecting the correct object when the labels differed by one consonant rather than when they differed by one vowel. Likewise, Escudero et al. (2016) found that Australian English-speaking adults were more successful at learning pseudowords that differed by one consonant (e.g., *bon* versus *ton*) in comparison to words that differed by one vowel (*deet* versus *dit*). These results suggest that the phonological encoding of consonants is stronger than the encoding of vowels, indicating that a consonant bias is present in lexical decision making with newly learned words.

2.6.5 Summary

Nespor et al.'s (2003) proposal that consonants are more beneficial than vowels in lexical processing tasks has been supported by extensive research across a wide range of experimental paradigms. Furthermore, these findings have been demonstrated in several different languages, suggesting a cross-linguistic advantage of consonants in word recognition and learning. From a developmental perspective, such research raise the question of whether infants are born with a consonant bias for lexical processing tasks and a vowel bias for syntactic and prosodic processing. Indeed, an innate knowledge of the functional roles of consonants and vowels would potentially assist infants in their acquisition of their native language.

2.7 Developmental origins of the consonant bias

Three distinct developmental hypotheses of the origins of a consonant bias have been presented. The initial bias hypothesis (Nespor et al., 2003) suggests that consonants and vowels are processed as singular linguistic features from birth, irrespective of the language they are exposed to. In contrast, the lexical hypothesis (Keidel et al., 2007) and the acoustic-phonetic hypothesis (Floccia et al., 2014) advocate different developmental trajectories and cross-linguistic differences based upon an infant's native language. The lexical hypothesis proposes that a consonant bias will be acquired once infants have developed an understanding of the lexical regularity of their native language between 12 and 24 months. Conversely, the acoustic-phonetic hypothesis advocates that such a bias will only be present following phonological acquisition at around 12 months.

2.7.1 Initial bias hypothesis

The "division of labour" between consonants and vowels, as proposed by Nespor et al. (2003), suggests that each phonetic category has a complementary functional bias; a consonant bias for the lexicon and a vowel bias for grammar. Consequently, Nespor et al. (2003) suggested that there could be an initial bias, present from birth, for infants to demonstrate an innate knowledge of the roles of consonants and vowels to benefit language learning. Accordingly, an initial consonant bias would assist infants in building the lexicon of their native language. In contrast, an initial vowel bias would provide infants with grammatical and prosodic information. Furthermore, if the differential processing of consonants and vowels is innate, then a consonant bias should be demonstrated in lexical tasks from birth, irrespective of language or the distributional and acoustic features of a language. Therefore, an initial bias would assume a limited impact of input characteristics, with neither developmental nor cross-linguistic differences found in the acquisition of a consonant bias.

2.7.2 *Lexical hypothesis*

The lexical hypothesis (Keidel et al., 2007) predicts that the contrasting roles of consonants and vowels arise from variations in the distribution of phonemes across languages and the extent to which they are advantageous in coding the lexicon. This hypothesis was proposed as an explanation to Bonatti et al.'s (2005) finding that French adults used consonants, rather than vowels, to segment words from an artificial speech stream. Keidel et al. (2007) suggested that the participants' acquired knowledge of their native lexicon, with consonants being more informative than vowels in recognising words in French, led to the preferential processing of consonants found in Bonatti et al. (2005). To demonstrate this, Keidel et al. (2007) analysed 4,943 CVCVCV words taken from the French corpus Lexique 3 (New et al., 2004), to identify the number of unique consonant and vowel tiers present in each word. For example, the French CVCVCV word *casino* contains the consonant tier *C-S-N* and the vowel tier *A-I-O*. The 4,943 CVCVCV words were found to contain 820 specific three consonant tiers and 562 specific three vowel tiers. This means that, on average, each unique consonant tier was found in 6.03 words, with each unique vowel tier found in 8.8 words. Therefore, the vowel tiers for CVCVCV words produced 1.46 times as many possible words compared to consonant tiers. Thus, even though consonants may outnumber vowels in French (17 consonants, 15 vowels), unique consonant tiers are heard less frequently than unique vowel tiers (Keidel et al., 2007; Nazzi & Cutler, 2019). Furthermore, an additional analysis showed that the CVCVCV words were approximately 40% more likely to be recognised based solely on their consonant tiers in comparison to their vowel tiers. Taken together, this demonstrates that consonants are more informative than vowels in identifying words.

In relation to the development of the division of labour between consonants and vowels, the lexical hypothesis proposes that the consonant bias is not innate. Instead, it suggests that the consonant bias will emerge once infants have reached an undetermined level of exposure to their native language, whereby the statistical informativeness of consonants, in comparison to vowels, in word recognition is learned. This would occur during the second year of life once infants' have acquired a sufficient lexicon and/or an understanding of the distribution of consonants and vowels in their lexicon, at which point they would use this knowledge to demonstrate privileged processing of consonants over vowels when identifying words. However, the size of an infant's lexicon or the amount of distributional information that is required for the bias to develop is unclear. Similarly, it is unknown whether infants would require an understanding of the meanings of words to develop a consonant bias or if it can be learned based on a recognition of the sound forms of words, which may impact when in development the bias would appear (Nazzi et al., 2016). Furthermore, the lexical hypothesis affords the possibility for cross-linguistic differences in the acquisition of lexical biases, due to variations in how informative the consonants and vowels of a given language may be in recognising words. For example, French-learning infants may demonstrate an early consonant bias due to how informative consonants are in comparison to vowels in word recognition. In contrast, infants learning languages where consonants are less informative, such as Danish, may not display a consonant bias until later in development.

2.7.3 Acoustic-phonetic hypothesis

The acoustic-phonetic hypothesis (Floccia et al., 2014) proposes that separate processing biases emerge due to the acoustic dissimilarities that exist between consonants and vowels. Specifically, consonants normally contain less energy, are shorter, and are therefore not as salient as vowels (Repp, 1984). Consonants tend to have large changes in frequency over time, whereas vowels have a steadier frequency (Liberman et al., 1967). Similarly, consonants are perceived categorically (Liberman et al., 1957), whereas vowels are perceived more continuously (Pisoni, 1973). Such acoustic dissimilarities should lead to the formation of two separate phonological categories, creating the functional asymmetry between consonants and vowels.

Developmentally, the acoustic-phonetic hypothesis suggests that a consonant bias will emerge by 12 months once infants have learned the phonological categories of their native language and developed a sensitivity to acoustic-phonological characteristics. Before this, infants may show an initial vowel bias in lexical processing. Infants' perception of vowel categories has narrowed to focus on native contrasts by 6 months (Kuhl et al., 1992; Polka & Werker, 1994), whereas an attunement towards native consonant contrasts occurs later at around 10 to 12 months (Best et al., 1988; Werker & Tees, 1984). This suggests that vocalic information is available earlier to infants than consonantal information. Correspondingly, the increased loudness and length of vowels compared to consonants may make them initially easier to process and more recognisable to infants (Mehler et al., 1978). Indeed, newborns can already distinguish between some native vowels (Moon et al., 2013). Furthermore, the pitch characteristics of infant-directed speech are also carried by vowels (Fernald & Kuhl, 1987). However, having been exposed to the acoustic and phonetic details of their native language, infants would learn the functional differences between phonetic categories, and switch to a consonant bias in lexical processing. For example, learning that consonants are processed more categorically than vowels (Fry et al., 1962) would highlight to infants that consonants are more dependable when both recognising and learning new words (Hochmann et al., 2011). Additionally, it may arise due to vowel categories appearing to be less discriminable than consonant categories (Bouchon et al., 2015). This would additionally coincide with the improved fine temporal resolution in infants' auditory systems by the end of the first year of life, allowing the accuracy needed to perceive the fast changes associated with consonant perception (Werner et al., 1992).

Furthermore, the acoustic and phonetic dissimilarities between languages leads to the possibility that cross-linguistic differences could emerge in the development of lexical processing biases. For instance, languages can differ in the ratio of consonants and vowels they contain. Similarly, although consonants typically outnumber vowels, there are systems with a higher vowel ratio. Other examples of variability between languages can include the presence of phonological short or long vowel contrasts and the degree of consonantal and/or vocalic reduction. Such variations could contribute to either strengthen or weaken infants' distinctions between the roles of consonants and vowels.

2.8 Evidence of lexical biases in infants

To investigate which hypothesis may account for the development of lexical processing biases, cross-linguistic research has looked at the role that consonants play in both word recognition and word learning tasks in infants. If a consonant bias is innate, as proposed by the initial bias hypothesis, then infants should show a preferential processing of consonants over vowels in lexical tasks from birth, irrespective of the lexical and acoustic/phonological features of their native language. However, evidence of a contrasting sensitivity to consonants and vowels over the course of development would provide support for either lexical or acoustic/phonetic properties modulating the acquisition of a consonant bias. Furthermore, the lexical hypothesis and the acoustic-phonetic hypothesis differentiate between when a consonant bias should first be found. The lexical hypothesis would predict that the bias would develop during an infant's second year alongside lexical development. In contrast, the acoustic-phonetic hypothesis anticipates the bias to emerge before infants have acquired a significant lexicon, instead estimating it to appear during the first year of life.

2.8.1 Lexical biases in French-learning infants

The initial developmental research into the acquisition of a consonant bias in lexical processing was conducted on French-learning toddlers. In the first such study, Nazzi (2005) measured 20-month-old French-learning infants' novel word learning using a name-based categorisation task (Nazzi & Gopnik, 2001). Firstly, a presentation phase showed infants three novel items. Two of the items were named with the same label (e.g., /pize/) by an experimenter, while the third was labelled with a one-feature phonetic change, either a stop consonant change (e.g., /tize/) or a vowel change (e.g., /pyze/). In a categorisation phase, the experimenter then held one of the identically named pair of objects and asked infants to provide them with the object that shared the same label from the two remaining objects. If infants displayed a consonant bias at this age, then it would be predicted that they would be successful in providing the correct object in consonant-contrasted pairs, but they would struggle to select the right object in vowel-contrasted pairs. Nazzi (2005) reported that infants were significantly better at choosing the correct object in the consonantally contrasted pairs but only performed at chance with vowel contrasting pairs. Thus, French-learning infants were demonstrating a consonant bias at 20 months, giving greater weight to consonant than vowel information in such a word learning task.

The initial finding by Nazzi (2005) of a consonant bias in word learning has been supported and furthered by a range of additional research in French-learning infants. For example, Nazzi and New (2007), again utilising the name-based categorisation task, discovered that 20-month-olds demonstrated a consonant bias for fricative (e.g., /fepɔd/ versus /ʃepɔd/), liquid (e.g., /rize/ versus /lize/) and nasal (e.g., /nuk/ versus /muk/) consonant contrasts. Alongside the stop consonant change demonstrated in Nazzi (2005), this suggests that the consonant bias in French-learning infants occurs irrespective of the type of consonant change made in word learning tasks. Additionally, there is evidence that the consonant bias is generalised across positions in a word. Nazzi and Bertoncini (2009) examined French-learning 20-month-olds' abilities to learn new words that contrasted either on their onset (e.g., /pɔd/–/bɔd/) or their coda (e.g., /pid/ versus /pit/) positions in a name-based categorisation task. They found that infants were successful in learning words in both positions. Thus, the consonant bias occurs irrespective of phoneme position in French-learning infants. Relatedly, Nazzi and Polka (2018) taught Canadian French-learning 20-month-olds pairs of words that differed on their initial vowel (e.g., /ɔpsi/ versus /œpsi/) or consonant (e.g., /upsa/ versus /utsa/), finding that performance was again better in the consonant condition. Therefore, consonant information was given more weight than vowel information in word learning, even when the initial phoneme of the target word was a vowel. Taken together, these studies show that, at 20 months, the consonant bias in word learning tasks is robust in French-learning infants, occurring irrespective of the type of consonant change, the position of the consonant change, and when in competition with word-initial vowels.

Research suggests that the consonant bias in word learning exists in both younger and older French-learning infants. Havy and Nazzi (2009) used a modified version of the name-based categorisation task to examine if 16-month-old French-learning infants also preferentially used consonant over vowel information when learning new words. Infants were first shown two novel objects, each labelled with a phonologically contrasted pseudoword that differed by either a one-feature consonantal contrast (e.g., /dul/ versus /gul/) or a one-feature vocalic contrast (e.g., /dɛr/ versus /dar/). The experimenter then produced a third novel object, naming it with one of the same labels as the previously seen objects, before placing it in a cup. Word learning was then tested by asking infants to put the identically named object in the cup as well. Using this simplified procedure, it was found that even 16-month-olds demonstrated a consonant bias, achieving the task when words differed in their consonants but performing at chance when words differed by their vowels. At 30 months, an age when infants are proficient at detecting vocalic information, infants demonstrate a consonant bias in such a word matching task (Nazzi et al., 2009). Similarly, 3-, 4-, and 5-year-olds are faster at detecting novel objects on a screen when their previously learned labels contrast with a distracter's label by one consonant as opposed to contrasting by one vowel (Havy et al., 2014). Thus, it appears that, for word learning tasks, French-learning infants display a consonant bias from around 16 months of age, which remains present throughout childhood and beyond.

Evidence from word recognition studies indicates that a consonant bias is present in French-learning infants. Zesiger and Jöhr (2011) used a preferential looking task to show French-learning 14-month-olds pictures of familiar objects. The images were presented in pairs on a screen. One of the objects was then correctly labelled or mispronounced, either with a consonant change or a vowel change. Additionally, the mispronunciations occurred either in the initial or final syllable of the word. Infants' looking time to the named object and the distractor image were measured during both pre- and post-naming, with greater looking times to the target once it had been labelled indicating word recognition. The results showed that infants' word recognition was blocked by consonant changes only in the second syllable, but vowel changes in either the initial or final syllable did not impact their ability to process the known words (though lack of recognition of familiar words in some conditions of this study weakens the evidence of a consonant bias). This finding suggests consonants had more influence in comparison to vowels in infants' recognition of familiar words.

Similarly, Poltrock and Nazzi (2015) explored whether 11-month-old French-learning infants listened longer to either consonant or vowel alterations of familiar word forms. In a baseline experiment, infants listened to lists of correctly pronounced familiar word forms (e.g., /wazo/, meaning *bird*) and lists of pseudowords (e.g., /walø/) in a head-turn preference task. As previously demonstrated in several studies (e.g., Hallé & de Boysson-Bardies, 1994, 1996; Vihman et al., 2004), they found that infants preferred listening to the lists of correctly pronounced word forms. In a second experiment, infants listened to lists of one-feature consonant (e.g., /wavo/) and one-feature vowel (e.g., /wazu/) mispronunciations of the same familiar word forms. Infants displayed longer listening times to the vowel mispronunciations in comparison to the consonant mispronunciations. Due to the previously found preference for familiar words, this result indicates that infants considered vowel mispronunciations to be more similar to correctly pronounced words compared to consonant mispronunciations. Additionally, Poltrock and Nazzi (2015) found that there was no relationship between infants' receptive vocabulary size and their listening preferences. These findings show that French-learning infants, even when they are only at the onset of lexical learning, already depend more on consonants than on vowels during word form recognition.

Research in even younger French-learning infants suggests that the consonant bias is not an innate feature of speech perception, but instead develops during an infant's first year. Bouchon et al. (2015), again using the head-turn preference paradigm, investigated whether French-learning 5-month-olds' own name recognition (a task they are proficient in at this age, Mandel et al., 1995) was impacted by one-feature consonant and one-feature vowel mispronunciations. Infants listened to alternating lists of their own name either correctly pronounced or with either phoneme changes on the initial consonant (e.g., *Victor* versus *Zictor*) or on the initial vowel (e.g., *Alex* versus *Elix*). The findings revealed that infants' listening times were comparable when hearing their correctly pronounced name and a consonant alteration. However, they attended longer to correctly pronounced name when it was paired with a vowel alteration. Thus, infants relied more on vocalic than consonant information in recognising their own name.

Furthermore, Bouchon et al. (2015) performed post-hoc acoustic analysis on the sound files presented to infants to assess how acoustic–phonetic properties impacted infants' responses. Three acoustic dimensions were measured, namely intensity, duration, and Mel-Frequency-Cepstrum Coefficients (MFCC). MFCC is a spectral based method of extracting features from voice signals and can supply a measurement of the acoustic distance found between two speech sounds. Although consonants were found to be shorter in length and softer in intensity compared to vowels, neither duration nor intensity were found to impact infants' preferences when listening to their own name. However, by comparing the MFCC features of the vowel changes, it was discovered that infants presented with a vowel mispronunciation in their name (e.g., *Alex* versus *Elix*) showed an increased preference for the correct pronunciation when there was a large acoustic distance between the two vowels (e.g., the distance between /a/ and /e/). Given that only one-feature changes were made to the vowels, the impact of acoustic distance occurred irrespective of phonetic distance. This indicates that French-learning 5-month-olds not only demonstrate a vowel bias in word recognition, but that this bias may be impacted by acoustic based metrics. Importantly, this result does not reinforce an initial bias hypothesis (Nespor et al., 2003), which proposes that a consonant bias should be observed at the beginning of lexical processing. Instead, this suggests that a consonant bias is learned.

Further research has assessed at what age infants may switch from an initial vowel bias, as demonstrated by Bouchon et al. (2015), to a consonant bias in lexical processing tasks. Nishibayashi and Nazzi (2016) examined this by using a head-turn preference task to explore French-learning 6- and 8-month-old infants' preferences for consonant and vowel mispronunciations of previously segmented words. Building on past research that had demonstrated that French-learning 6- and 8-month-olds are adept at segmenting monosyllabic words contained within fluent speech (Goyet et al., 2013; Nishibayashi et al., 2015), infants first attended to two short stories that included two repeated monosyllabic target words. Infants' listening times to one-feature consonant and vowel changes made to the target words were then measured. At 6 months, infants listened longer to the vowel mispronunciations, suggesting that word recognition was more impacted by consonant changes. This result suggested that French-learning infants initially demonstrate a vowel bias, replicating the findings in Bouchon et al. (2015). However, Nishibayashi and Nazzi (2016) found the opposite effect at 8 months, with infants listening longer to the consonant mispronunciations. Such a consonant bias was found when the phoneme change was made to either the onset or coda position, and across several different phoneme contrasts. This finding is further supported by an ERP study, which reported that 8-month-olds show a consonant bias in a comparable segmentation task (Von Holzen et al., 2018). Therefore, a consonant bias is found in French-learning infants during word segmentation by the time they are 8-months-old.

A more recent study by Von Holzen and Nazzi (2020) demonstrated that French-learning infants switch from a vowel bias to a consonant bias during their first 12 months. This experiment used the head-turn preference procedure to first familiarise 5-, 8-, and 11-month-old French-learning infants with correctly pronounced tokens of their own name, before measuring listening times towards one-feature consonant and one-feature vowel alterations to their name. The findings showed the opposite pattern to previous research at these ages (Bouchon et al., 2015; Nishibayashi & Nazzi, 2016; Poltrock &

Nazzi, 2015). Both 5- and 8-month-old infants preferred vowel mispronunciations of their own name, whereas 11-month-olds preferred consonant mispronunciations. This finding was interpreted as infants showing a novelty preference in this task, rather than favouring familiar sounding (Bouchon et al., 2015; Poltrock & Nazzi, 2015) or previously familiarised (Nishibayashi & Nazzi, 2016) word forms. Such a novelty effect suggests that infants showed a vowel bias at 5- and 8 months, with a consonant bias appearing at 11 months.

Von Holzen and Nazzi (2020) suggested that differences in their experimental design, compared to previous research, may explain why infants displayed a novelty effect as opposed to a familiarity effect. Infants often display a novelty preference when an experimental task has low difficulty or long familiarisation times (Hunter & Ames, 1988). Although task difficulty was similar between Von Holzen and Nazzi (2020) and previous consonant bias research at this age, the degree of familiarisation differed. Infants received a longer familiarisation, listening to correct pronunciations of their own name, compared to experiments in which infants had no familiarisation at all to the target words (Bouchon et al., 2015; Poltrock & Nazzi, 2015) or were familiarised with a novel word embedded within a passage (Nishibayashi & Nazzi, 2016). Therefore, such a familiarisation phase may have led infants to prefer novel sounding mispronunciations over familiar sounding alterations.

The results in Von Holzen and Nazzi (2020) differed from previous research into the timing of the switch from a vowel to a consonant bias. Nishibayashi and Nazzi (2016) found that French-learning 8-month-olds displayed a consonant bias for previously segmented words. In contrast, Von Holzen and Nazzi (2020) found that infants maintained a vowel bias at this age in a word form recognition task, only switching to a consonant bias at 11 months. This later timing of the switch in biases may be explained by differences in the cognitive processes required in the tasks presented to infants. The segmentation task in Nishibayashi and Nazzi (2016) required short-term recall of recently heard novel words. In comparison, infants retrieved a previously heard word in the word recognition task used by Von Holzen and Nazzi (2020). Therefore, the complexity of the task presented to infants may impact whether they display a consonant or vowel bias during their first year of life.

Furthermore, Von Holzen and Nazzi (2020) measured how acoustic factors impacted infants' preferences. An analysis of the intensity, duration, and the MFCC features of the stimuli presented to infants found no significant effect on infant listening times. This contrasts with the increased preference for greater acoustic differences between vowels as reported by Bouchon et al. (2015). However, both Von Holzen and Nazzi (2020) and Bouchon et al. (2015) analysed the acoustic factors post-hoc, rather than explicitly manipulating them during recording of the stimuli, which may have resulted in the disparity of findings.

Finally, Von Holzen and Nazzi (2020) assessed how lexical factors impacted 8- and 11-month-old infants' listening times towards consonant or vowel mispronunciations. Firstly, they calculated the number of words infants knew, as measured by the French Communicative Developmental Inventory: Words and Gestures for ages 8 to 16 months (Kern & Géraldine, 2003). From these words, phonetic transcriptions were then used to create separate consonant and vowel tier proportion scores. For instance, if an infant understood the words *bain* (/bɛ̃/, meaning *bath*), *chien* (/ʃjɛ̃/, meaning *dog*), and *merci* (/mɛʁsi/, meaning *thank you*), then they would know three distinct consonant tiers, as each word has a unique consonant sequences (/b/; /ʃ.j/; /m.ʁ.s/). However, they would only know two distinct vowel tiers, with *bain* and *chien* containing the same vowel sequence (/ɛ̃/), which is distinctive from the vowel sequence in *merci* (/ɛ.i/). To create an individual infant's proportion score, the number of unique consonant tiers and vowel tiers they knew was divided by how many words they knew. The results found that both lexicon size and proportion of consonant and vowel tiers known did not impact infants' listening times, even though infants were found to know more unique consonant pairings than vowel pairings. This added further support to existing research (Poltrock & Nazzi, 2015) that lexical factors do not appear to influence the presence of a consonant bias at this early age in French-learning infants.

In summary, research in French-learning infants suggests that a consonant bias in lexical processing is present from an early age. From 16 months, infants depend less on vowels and more on consonants when learning new words (Havy & Nazzi, 2009; Nazzi, 2005), a finding that extends into childhood and beyond (Havy et al., 2014), and one that is unrelated to the category (Nazzi & New, 2007) or position (Nazzi & Bertoncini, 2009) of the consonant change. However, rather than being an innate bias, French-learning infants appear to first show a reliance on vowels in lexical tasks (Bouchon et al., 2015), before switching to a consonant bias between the ages of 8 (Nishibayashi & Nazzi, 2016) and 11 months (Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020).

2.8.2 Lexical biases in English-learning infants

Although evidence of a consonant bias appears to be robust in French-learning infants from the first year of life onwards, evidence from infants learning other languages suggests that cross-linguistic differences may exist in the acquisition of lexical processing biases. Firstly, the developmental course of the consonant bias in English-learning infants is debatable. French and English differ on several phonological and acoustic factors. The ratio of consonants and vowels vary between French (17 consonants, 15 vowels) and English (24 consonants, 12 vowels). In relation to consonants, American English contains more fricatives and affricates, as well as one more nasal, compared to French (Delattre, 1964). There are also significant differences between the French and English vowel systems. For example, French vowels are typically similar in duration, are high, rounded, and fronted, and contain reduced diphthongisation. In contrast, American English vowels typically vary in duration, are low, unrounded, and back, with the majority diphthongised (Delattre, 1964). This differentiation is comparable between French and British English, with British English vowels moderately diphthongised in a several accents (Giegrich, 1992). Similarly, both British English (White et al., 2012) and American English (Delattre, 1966) have more vowel reduction than French. English syllables are characteristically heavier than French in relation to their number of consonants, resulting in a longer duration and larger variability of consonant clusters in English (Ramus et al., 1999; White et al., 2012). Relatedly, French has fewer monosyllabic words compared to English (Delattre, 1964).

Additionally, English is a stress-timed language. Acoustically, stress can impact the duration and intensity of vowels (Fear et al., 1995). As a syllable timed language, French has no lexical stress (Hirst et al., 2001). However, in phrase-final positions, content words' final syllables are typically lengthened in French (Delattre, 1966). Developmentally, English-learning infants can differentiate between stress-initial and stress-final words by the time they are 9 months (Jusczyk, Goodman, et al., 1999). In comparison, French-learning infants can only discriminate stress patterns following a long period of familiarisation (Bijeljac-Babic et al., 2012; Skoruppa et al., 2009).

The dissimilarities between French and English suggest that developmental differences may exist in the acquisition of a consonant bias between the two languages. The acoustic differences in vowels, increased consonant clusters, and variations in stress may result in English-learning infants, in comparison to French-learning infants, facing a more complex task when acquiring their native phonetic and phonemic categories. This may result in English-learning infants demonstrating difficulties in detecting consonant and vowel changes in lexical tasks.

Data from word learning, word recognition, and word form recognition studies indicate that a consistent consonant bias does not emerge until later in development in British English-learning infants compared to French-learning infants. Nazzi et al. (2009) found that consonant information is given more weight than vowel information in name-based categorisation word learning tasks in both French-learning and British English-learning 30-month-olds. However, research in younger British English-learning infants reveals a different pattern. Floccia et al. (2014) tested British English-learning 16-, 23-, and 30-month-olds' reliance on consonants in word learning using the same simplified version of name-based categorisation task as Havy and Nazzi (2009). Whilst 30-month-olds performed better when learning consonant contrasting word pairings, infants at 16 and 23 months did not show such a consonant bias. Instead, infants learned consonant and vowel contrasting word pairings equally well. This finding contrasts with French-learning infants (e.g., Havy & Nazzi, 2009; Nazzi, 2005; Nazzi & Bertoncini, 2009), who display a clear consonant bias in name-based categorisation tasks at these ages. This suggests that, not only is the consonant bias in lexical processing not innate, but language specific mechanisms may impact the acquisition of the bias.

Cross-linguistic variations between French-learning and British English-learning infants have been found in word recognition tasks. Mani and Plunkett (2007) looked at how vowel and consonant mispronunciations impacted British English-learning 15-, 18-, and 24-month-olds' word recognition. Using an IPL procedure**,** infants were shown pairings of familiar objects on a screen that were either labelled correctly (e.g., *ball*; /bɔːl/) or with a word-medial vowel change (e.g., *bule;* /buːl/). Word recognition was evaluated by measuring infants' looking times to the target image. Infants in all three age groups looked longer at a target referent when it was correctly labelled but not when it was named with a vowel mispronunciation, indicating that vowel alterations impaired word recognition. In a second study, 15-, 18-, and 24-month-olds were again presented with the same task, but with additional word-initial consonant mispronunciations (e.g., *gall*; /gɔːl/), alongside the correct and vowel mispronunciations. Here, 15-month-olds' word recognition was impacted by consonant alterations but not vowel alterations, suggesting a consonant bias at this age. In contrast, 18- and 24-month-olds were equally sensitive to consonant and vowel mispronunciations, failing to attend to the target when either phoneme was altered. Additionally, no evidence was found for a relationship between vocabulary size and sensitivity to mispronunciations. Similarly, using the same method, Mani and Plunkett (2010) found that British English-learning 12-month-olds' lexical access appears to be equally impacted by consonants and vowels. However, infants with larger vocabularies were more sensitive to vowel alterations, but not consonant alterations, in comparison to infants with smaller vocabulary sizes. Thus, the findings in British English-learning infants differentiate from the consistently found consonant bias in French-learning infants in word recognition tasks (e.g., Poltrock & Nazzi, 2015; Zesiger & Jöhr, 2011), adding further evidence that cross-linguistic differences exist in the acquisition of a consonant bias.

The development of the consonant bias in word form recognition during British English-learning infants' first year of life is an under-researched area. In the only study to date, Delle Luche et al. (2017) examined whether British English-learning 5-month-olds show a vowel bias when listening to mispronunciations of their own name. Using the head-turn preference procedure, a first experiment presented infants with lists of their own name being correctly pronounced or an altered version with either a one-feature onset consonant change (e.g., *Daniel* versus *Taniel)* or a one-feature onset vowel change (e.g., *Esme* versus *Usme*). The results showed that, rather than an initial vowel bias as demonstrated in French-learning infants (Bouchon et al., 2015; Nishibayashi & Nazzi, 2016; Von Holzen & Nazzi, 2020), British English-learning infants did not identify their own name in comparison to either phoneme change. To explore whether such a result could be explained by British English-learning infants failing to detect their own name at this early age, Delle Luche et al. (2017) conducted a second experiment, whereby an infant's name (e.g., *Ella*) was paired with a phonetically different name, which shared the same syllables and stress patterns (e.g., *Robyn*). When presented with this task, infants displayed a preference for their own name. This suggests that the absence of a consonant bias was not due to British English-learning infants being delayed in recognising their own name. Instead, the lack of a lexical bias suggests that, at 5 months, British English-learning infants' early word representation does not contain phonetic details.

Furthermore, post-hoc analysis by Delle Luche et al. (2017) on how acoustic factors of the stimuli presented to infants impacted their preferences also revealed cross-linguistic differences. In the first experiment, when presented with a consonant change, British English-learning infants were more likely to listen longer to the correct pronunciation of their name if their name started with a short amplitude-rise-time consonant (stops, e.g., *Poppy*) in comparison to those whose name began with a longer amplitude-rise-time consonant (continuants, e.g., *Finn*). This result was found irrespective of whether the phoneme in the mispronunciation involved a stop (e.g., *Poppy* versus *Toppy*) or a continuant (e.g., *Ben* versus *Wen*) consonant, indicating that infants' own name recognition was more reliable when beginning with an onset stop. Energy can be used as a cue to differentiate between consonants and vowels, stops and continuants (Stevens & Blumstein, 1981), as well as stress patterns (Jusczyk et al., 1993). Energy has been shown to impact infants' proficiency at discriminating and extracting speech sounds, with Polka et al. (2001) attributing the inability to distinguish between some phonemes (e.g., the /d/-/ð/ contrast) at 10 to 12 months to the low energy they exhibit. This suggests that 5-month-old British English-learning infants did not rely on detailed phonetic features in word recognition but were instead using broadly defined phonetic features (i.e., stops versus continuants) and acoustic cues to identify their name. Similarly, they were using different acoustic cues in comparison to French-learning infants, who used the acoustic distance in vowels to differentiate phonetic changes (Bouchon et al., 2015).

Research in other dialects of the English language provide further evidence of cross-linguistic differences in the consonant bias acquisition. American English-learning infants under the age of 17 months can struggle in word-learning tasks when they are required to learn two word-object pairings that are contrasted by only one consonant. For example, Stager and Werker (1997) found that American English-learning 14-month-olds failed to learn two novel words when they were phonetically similar (e.g., *bih* and *dih*) but were able to learn them when they were phonetically distinct (e.g., *nim* and *leef*). However, 14-month-olds could discriminate between the similar sounding words in tasks that did not require word learning. Infants also failed to detect mismatches when presented with minimal consonants (Pater et al., 2004). In relation to vowels, 15-month-old Canadian English-learning infants can learn words that contrast in vowel height (e.g., *deet* and *dit*) but not vowel backness (e.g., *deet* and *doot*), indicating that salient acoustic cues for vowels are more beneficial for word learning at this age (Curtin et al., 2009). Similarly, 15-month-old Australian English-learning infants can detect changes in vowel minimal pairs in a word learning task when the acoustic difference between the two contrasts is large (Escudero et al., 2014). Finally, Johnson et al. (2003) found that 12-month-old American English-learning infants demonstrated an adult-like bias in a word segmentation task. Specifically, when familiarised with words such as *win* or *low*, infants recognised *window* or *below* but not *wind* or *slow*. Infants considered a solitary consonant (e.g., *s* in *slow*) as a segment of the ongoing lexical frame. In contrast, a CV (e.g., *be* in *below*) was consisted as a separate frame, allowing for the familiarised word to be detected. Therefore, as well as cross-linguistic differences in the acquisition of the consonant bias, there may be within-linguistic differences, with dialect impacting consonant and vowel perceptual development.

Taken together, the acquisition of a consonant bias in lexical processing in English-learning infants appears to follow a different developmental course. British English-learning infants display a consonant bias in word learning tasks at 30 months (Nazzi et al., 2009) but learn consonant and vowel contrasting word pairings equally well at 16 and 23 months (Floccia et al., 2014). In word recognition studies, only 15-month-olds show a consonant bias, with 18- and 24-month-olds (Mani & Plunkett, 2007), as well as 12-month-olds (Mani & Plunkett, 2010), showing equal sensitivity to consonants and vowels. At the beginning of lexical acquisition, British English-learning 5-month-olds do not appear to show any lexical biases but may rely on the energy in consonants in a word form recognition task (Delle Luche et al., 2017). Furthermore, the developmental course of lexical processing biases may differ within categories of the same language (Johnson et al., 2003).

2.8.3 Lexical biases in Italian-learning infants

Data from Italian-learning infants suggests that the developmental trajectory of the consonant bias follows a similar pattern to French-learning infants. Italian is a language that is both lexically and rhythmically similar to French. Italian-learning newborns appear to have a better memory for vowel information in comparison to consonantal information in their word form recognition. For example, Benavides-Varela et al. (2012) first familiarised Italian-learning newborns, with an average age of 2.5 days, with a CVCV pseudoword (e.g., /mimi/). In a test phase, half of the infants listened to a word that preserved the familiarised word’s consonants (e.g., /momo/) and the other half listened to a word that preserved the familiarised word’s vowels (e.g., /sisi/). fNIRS was used measure whether the newborns could detect phonetic change. fNIRS is a non-invasive brain imaging method which uses multiple sensors to detect changes in oxyhemoglobin and deoxyhemoglobin caused by cortical activation (Quaresima et al., 2012). The results revealed contrasting blood flow patterns in newborns based upon the phonetic change they heard. Specifically, an increase in oxyhemoglobin was found in newborns presented with a test word that contained novel vowels (e.g., /mimi/ - /momo/), whereas a decrease in oxyhemoglobin was found in newborns presented with a test word containing novel consonants (e.g., /mimi/ - /sisi/). This effect was strongest in the newborns’ right frontal areas, which correspond to regions of an adult’s brain activated when recalling verbal information (e.g., Rugg et al., 1996; Shallice et al., 1994). This finding suggests that Italian-learning newborns have a vowel bias, rather than a consonant bias, in lexical processing tasks from birth.

Research in older Italian-learning infants has revealed a similar developmental trajectory for the acquisition of the consonant bias as found in French-learning infants. Hochmann et al. (2011) first familiarised Italian-learning 12-month-olds with two pseudowords that each predicted the appearance of objects on a screen. For example, *dudu* preceded a toy appearing on the left and *keke* preceded a different toy appearing on the right. In a test phase, infants heard an ambiguous word, formed from the consonants of one pseudoword and the vowels of the other (e.g., *dede*), or the opposite (e.g., /kuku/) in the absence of any object appearing on the screen. If infants had a consonant bias in word learning, they should look towards the location that retained the consonants of the familiarised pseudoword rather than the location that retained the vowels of a familiarised pseudoword. The findings showed that infants did indeed rely more on consonants, searching for the object in the location of the screen predicted by the consonants on the ambiguous pseudoword. More recently, Hochmann et al. (2017), using a similar method as Hochmann et al. (2011), found that Italian-learning 6-month-olds utilised vowels rather than consonants when looking for a previously labelled object. For example, they attended to the side of a screen previously connected with the pseudoword *dede* as opposed to the side previously connected with the pseudoword *kuku* when hearing the ambiguous pseudoword *keke*. Thus, as in French-learning infants, Italian-learning infants appear to demonstrate an initial vowel bias before switching to a consonant bias by their first birthday.

2.8.4 Lexical biases in Danish-learning infants

Danish provides an interesting comparison to the research discussed thus far as it is a language that contains more vowels than consonants. This in turn could lead to vowels being more informative than consonants in lexical tasks. Similarly, consonants in Danish are considerably reduced or underarticulated, resulting in an increased salience of vowels (Bleses et al., 2011). From both a lexical and acoustic-phonetic perspective, the increased number and salience of vowels in Danish should lead Danish-learning infants to favour vowels over consonants in lexical processing tasks. To test this, Højen and Nazzi (2016) presented Danish-learning 20-month-old infants with a name-based categorisation task, using similar pseudowords and objects as in Nazzi (2005). Danish-learning infants were able to learn pairs of words that differed by a vowel but failed to learn new words when they were contrasted by a consonant. Thus, Danish infants were demonstrating a vocalic bias in a word learning task, the opposite pattern found in French-learning infants at 20 months (e.g., Nazzi, 2005). Although adult data is not yet available to establish whether Danish adults also show a vocalic bias in lexical processing tasks, the findings from Danish-learning infants provide further evidence that the development of the consonant bias differs across languages. Furthermore, they demonstrate that the division of labour between consonants and vowels may not be universal.

2.8.5 Lexical biases in infants learning tone languages

The acquisition of lexical processing biases has recently been investigated in infants learning a tone language, such as Mandarin and Cantonese. Forming an estimated 70% of all languages (Yip, 2002), tone languages contain both consonants and vowels, as well as lexical tones, which signify syllable level changes that can be used to differentiate meaning at the word level. For example, the syllable /ji/ in Cantonese can mean *chair*, *idea*, *doctor*, *ear*, *child*, or *two* because of the tone used to produce it (Gómez et al., 2018). Lexical tones are primarily identified by changes in fundamental frequency cues, such as pitch height and contour, as well as other acoustic cues including duration and amplitude (Gandour, 1983). In relation to the development of lexical processing biases, tones could potentially strengthen a reliance on consonants, as they have a variable vowel realisation which may cause an increased difficulty in identifying vowels. Alternatively, tones could decrease or reverse a consonant bias since vowels carry tone information. Similarly, word structures in tone languages typically have a reduced consonant to vowel ratio, due to a constrained syllable phonology, which could also lead to a reduced advantage of consonants in lexical processing (Nazzi & Cutler, 2019). Indeed, there is recent evidence to suggest that Mandarin-speaking adults may show an opposite vowel bias in lexical processing tasks (Wiener, 2020).

Findings from Mandarin-learning children suggest the sensitivity to vowel, consonant, and tone alterations changes during development. In a preferential looking task, Singh et al. (2015) found that bilingual Mandarin-English-learning 2.5 to 3.5-year-olds were more sensitive to tone mispronunciations of familiar words compared to vowel or consonant mispronunciations. In contrast, 4- to 5-year-olds demonstrated an increased sensitivity to consonant and vowel variations over tone variations. However, both age groups appeared equally sensitive to consonant and vowel mispronunciations. Using the same task, Wewalaarachchi et al. (2017) found that 24-month-old Mandarin-learning monolinguals and Mandarin-English-learning bilinguals were comparably sensitive to consonant, vowel, and tone mispronunciations, again showing no consonant bias at this age. An analysis of the time course of target fixation revealed that infants differed in their response speed. Monolingual infants showed the highest sensitivity for tone, followed by vowel, and then consonant mispronunciations. Bilingual infants were more sensitive to vowel, followed by consonant, and then tone mispronunciations. In 6-year-olds, Wewalaarachchi and Singh (2020) found that monolingual Mandarin-learning and bilingual Mandarin-English-learning children were least sensitive to tone mispronunciations, with no difference in sensitivity between consonants and vowels, in both preferential looking and an explicit judgment task. Together, these findings suggest that lexical tones do not strengthen the development of a consonant bias, nor do they demonstrate an adult-like vowel-bias. Instead, Mandarin-learning infants display changes in their relative sensitivity to consonants, vowels, and tones over the first few years of life.

Finally, a recent study by Chen et al. (2021) assessed the consonant bias in infants learning Cantonese, a language that has a more complicated tone system than Mandarin. The identification of Cantonese tones requires both register and contour, while only contour is needed to identify Mandarin tones. Likewise, Cantonese has six tones, whereas Mandarin has four tones. They presented Cantonese-learning 20- and 30-month-olds with cartoons which named two novel objects. The pairs of objects differed by either their consonant (e.g., /pi3/ - /ti3/), vowel (e.g., /khim3/ - /khɛm3/), or tone (e.g., /khim1/ - /khim3/). In a test phase, the objects appeared side by side and were labelled. Eye-tracking was used to measure whether infants looked to the correct object. While 20-month-old infants did not learn the words in this task, 30-month-olds' were found to only learn the object names in the vowel condition. Thus, Cantonese-learning 30-month-olds were demonstrating a vowel bias in their word learning. This again suggests that the language an infant is exposed to impacts when a consonant bias is acquired.

2.8.6 Summary

Figure 2 displays a timeline of the research examining the role consonants and vowels play in lexical tasks in newborns to 30-month-olds. The findings of these studies have consistently shown that the privileged role of consonants, as found in word learning and word recognition in adults, emerges during development. Therefore, infants must learn the specialised function of consonants in lexical processing. Furthermore, such studies have demonstrated that the age in which a consonant bias is present varies cross-linguistically. This demonstrates that lexical and/or acoustic-phonetic properties of an infant's native language may shape the acquisition of the consonant bias.

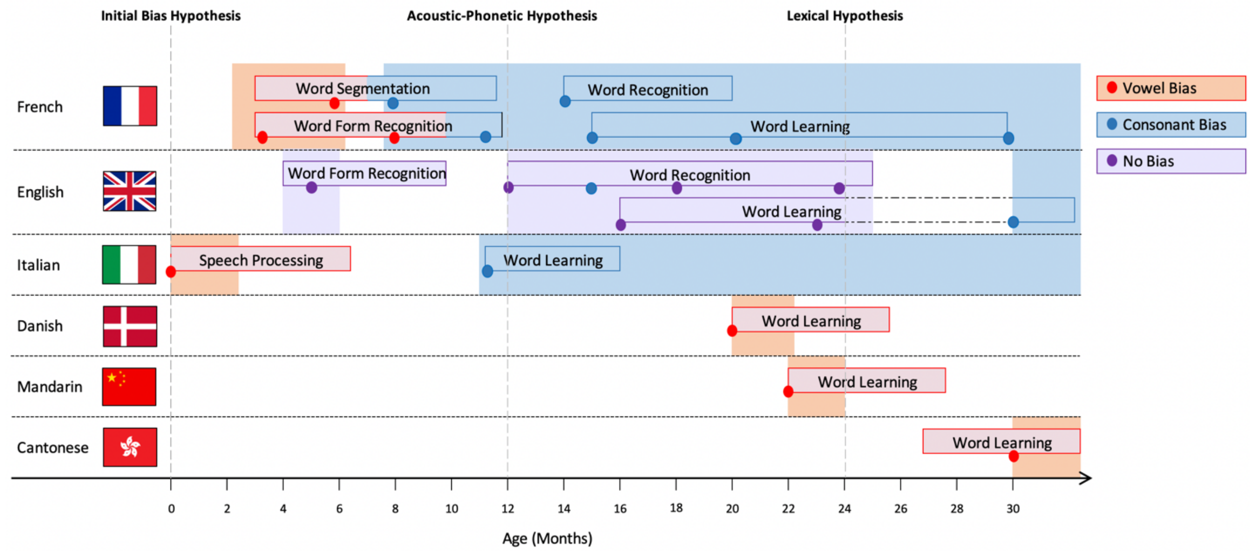


Figure 2: *Visual timeline of research reporting consonant and vowel processing in lexical tasks in infants from birth to 30 months, grouped by infants' native language. Each colour coded dot indicates the bias reported. Vertical horizontal lines represent the age in which each developmental hypothesis predicts the age of acquisition of the consonant bias* (*adapted from Nazzi et al., 2016).*

In relation to the initial bias (Nespor et al., 2003), lexical (Keidel et al., 2007) and acoustic-phonetic (Floccia et al., 2014) hypotheses, and the cross-linguistic differences between infants learning French (Bouchon et al., 2015; Nishibayashi & Nazzi, 2016; Von Holzen & Nazzi, 2020) and British English (Delle Luche et al., 2017; Floccia et al., 2014), it appears clear that the preferential role of consonants in lexical processing is not innate. Moreover, evidence suggests that infants may even begin with an initial vowel bias in lexical tasks (Benavides-Varela et al., 2012; Bouchon et al., 2015; Nishibayashi & Nazzi, 2016; Von Holzen & Nazzi, 2020). Thus, the initial bias hypothesis is not supported by the current developmental findings. On the other hand, the present data does not yet allow for the conclusion that the consonant bias develops because of either the lexical regularities or the acoustic-phonetic information found in an infant's native language. Indeed, both hypotheses may partially explain the development of lexical processing biases.

The lexical hypothesis proposes that how informative a language's phonemes are in coding the lexicon, as well as lexical acquisition, may lead to the preferential role of consonants in lexical tasks. This may explain why Danish infants display a later vowel bias (Højen & Nazzi, 2016) given that vowels outnumber consonants in this language and may therefore be more informative to infants. However, evidence of a consonant bias in a word segmentation task in French-learning 8-month-olds (Nishibayashi & Nazzi, 2016) challenges this claim. Such infants are unlikely to have either a large enough lexicon or sufficient linguistic experience with the distributional information provided by their lexicon to know that consonants are more informative compared to vowels for word recognition. Likewise, no link has been found at 11 months between the preference for consonants and either vocabulary size (Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020), or the proportion of consonant and vowel tiers known (Von Holzen & Nazzi, 2020). Similarly, the equal preference for consonants and vowels in British English-learning infants at 23 months (Floccia et al., 2014), when infants are likely to have acquired both a sizeable lexicon and an understanding of the phonological structure of their lexicon, also suggests that the lexical hypothesis alone cannot explain the development of the consonant bias. Nevertheless, it remains unknown how such lexical factors may impact younger British

English-learning infants' reliance on consonants and vowels in word form recognition tasks during their first year of life.

In contrast, the acoustic-phonetic hypothesis suggests that the consonant bias is linked to the early processing of acoustic/phonetic information and to the phonological properties of an infant's native language. This would attribute the early vowel bias found in French- (e.g., Bouchon et al., 2015) and Italian-learning (Benavides-Varela et al., 2012) infants to the earlier acquisition of native vowel categories (e.g., Kuhl et al., 1992) and the acoustic factors that make vowels initially easier to process (Mehler et al., 1978). It also explains their switch to a consonant bias by 12 months (e.g., Poltrock & Nazzi, 2015), with the acquisition of native consonant categories (Werker & Tees, 1984) and further exposure to their native language contributing to learning the functional differences between phonemes. This hypothesis may account for the cross-linguistic differences between French- and British English-learning infants, given the phonological differences between their two languages. Such dissimilarities may mean that British English-learning infants have a more complicated task in acquiring their phonetic categories, compared to French-learning infants, which may in turn lead to their difficulties in detecting phonetic changes in lexical tasks (e.g., Delle Luche et al., 2017; Floccia et al., 2014). Nevertheless, it is unclear what acoustic and phonetic information British English-learning infants process when hearing familiar word forms. Indeed, post-hoc findings in British English-learning 5-month-olds suggest they may use intensity rather than phonetic information to recognise word forms (Delle Luche et al., 2017).

To gain further understanding, additional cross-linguistic tests of the consonant bias at the onset of phonetic acquisition and of its potential links to acoustic-phonetic and lexical factors are required. Chapter Two will present a series of head-turn experiments that will focus on how consonant and vowel changes impact British English-learning infants' word form recognition during the first year of life. Firstly, it will be investigated whether 5-month-olds' familiar word form recognition is impacted by phonetic mispronunciations and energy alterations (Experiment 1). Rather than presenting infants with their own name, as in previous experiments, all infants will listen to variations of the same stimuli. Specifically, the familiar word form *mummy* will be heard in competition with either phonetic or intensity changes. The next experiments will then focus on the phonetic specificity of 11- and 12-month-olds' familiar word form recognition. In addition, these studies will explore whether lexical factors impact infants' preferences. It will first be established whether 11-month-old infants prefer listening to lists of correctly pronounced familiar words over pseudowords (Experiment 2a). Following this, 11-month-olds' preferences for onset consonant mispronunciations (Experiment 2b) and medial consonant mispronunciations (Experiment 2c) of the familiar words, each compared to vowel pronunciations, will be examined. Finally, 12-month-olds' recognition for a single correctly pronounced word, versus either a consonant or vowel mispronunciation, will be tested (Experiment 3).

3 Chapter Two

Note. Experiments 2a, 2b, and 3 are an adaptation of a paper: Ratnage, P., Nazzi, T., & Floccia, C. (2023). Vowels and consonants matter equally to British English-learning 11-month-olds' familiar word form recognition. Journal of Child Language, 1-24. https://doi.org/10.1017/s0305000923000223

3.1 Experiment 1: Recognition of a familiar word form following phonetic or intensity changes in British English-learning 5-month-olds

3.1.1 Introduction

Experiment 1 aimed to further investigate the developmental origins of the consonant bias during the first year of life by exploring how phonetic changes impacted British English-learning 5-month-old infants' word form recognition and if infants' recognition of correct pronunciations were modulated by alterations in intensity. This age was selected as infants are only just beginning to acquire their native lexicon, recognising some familiar word forms (Bortfeld et al., 2005; Mandel et al., 1995; Parise et al., 2010) and demonstrating comprehension of a few frequently heard words (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012). Similarly, infants have become attuned to the vowels of their native language by 6 months (e.g., Polka & Werker, 1994), whereas consonant acquisition occurs later at 10 to 12 months (e.g., Werker & Tees, 1984).

Research in French-learning infants suggests that, rather than being an innate bias as proposed by Nespor et al. (2003), a consonant advantage is not present until 11 months in word form recognition tasks (Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020) and 8 months when identifying segmented word forms (Nishibayashi & Nazzi, 2016). Before this, French-learning infants initially display a vowel bias when detecting a one-feature consonant or vowel change to their own name at both 5 and 8 months (Bouchon et al., 2015; Von Holzen & Nazzi, 2020), as well as in word form segmentation at 6 months (Nishibayashi & Nazzi, 2016). This early vowel advantage is also found in Italian infants, who display a vowel bias at birth (Benavides-Varela et al., 2012) and at 6 months (Hochmann et al., 2017), before showing a consonant bias at 12 months (Hochmann et al., 2011). Thus, a developmental shift from a vowel bias to a consonant bias occurs during the first year of life in French- and Italian-learning infants.

To date, only one study has examined the development of the consonant bias in British English-learning infants before their first birthday. Delle Luche at al. (2016) found that 5-month-olds did not detect their own name when contrasted with either a one-feature onset consonant or vowel alteration. This lack of a consonant or vowel bias in own name recognition in British English contrasts with the reliance on vocalic information in similar experiments in French (Bouchon et al., 2015; Von Holzen & Nazzi, 2020). Indeed, research in older British English-learning infants suggests that a consonant bias is not consistently found in lexical tasks until 30 months of age (e.g., Floccia et al., 2014; Nazzi et al., 2009). Differences in the presence of a consonant bias during infancy have also been found in Danish (Højen & Nazzi, 2016) and tone languages (Chen et al., 2021).

Nevertheless, relatively little is known about the mechanisms that cause cross-linguistic differences in the acquisition of the consonant bias. From a lexical perspective (Keidel et al., 2007), 5-month-olds may not yet have been exposed to a significant amount of their native language whereby the effectiveness of consonants, compared to vowels, has become clear. However, from an acoustic-phonetic perspective (Floccia et al., 2014), 5-month-olds may show an early vowel bias due to the increased saliency of vowels in comparison to consonants. Similarly, they are likely to have a more sophisticated comprehension of their native vowels compared to consonants at this early age (Kuhl et al., 1992; Werker & Tees, 1984).

The role of acoustic/phonetic factors in the development of the consonant bias has been investigated by analysing how duration, intensity, and spectral distance modulate infants' mispronunciation detection. In French-learning 5-month-olds, Bouchon et al. (2015) found that infants hearing a vowel mispronunciation in their own name showed an increased preference for a correct pronunciation when there was a large acoustic distance between the two vowels (e.g., the distance between /a/ and /e/). However, such acoustic features did not impact infants' own name preferences at 5, 8, or 11 months in Von Holzen and Nazzi (2020). Findings in Delle Luche et al. (2017) from British English-learning 5-month-olds suggest that, when presented with a consonant change to their own name, infants were more likely to listen longer to a correct pronunciation if their name started with a short amplitude-rise-time consonant (stops, e.g., *Poppy*) in comparison to those who name began with a longer amplitude-rise-time consonant (continuant, e.g., *Finn*). This suggested that British English-learning infants' word form recognition may be modulated by energy cues.

However, the own name studies conducted in French (Bouchon et al., 2015; Von Holzen & Nazzi, 2020) and British English (Delle Luche et al., 2017) did not control for the acoustic features of the stimuli presented to infants, instead analysing their properties post-hoc. As proposed by Von Holzen and Nazzi (2020), a different approach would be to directly manipulate the acoustic characteristic of the stimuli before measuring how such changes impact infants' listening times. Furthermore, providing each infant with alterations of the same familiar sounding word form, rather than their own name, would allow for each infant to be presented with a consistent phonetic feature change, and allow for a more controlled measure of infants' preferences.

Experiment 1 adopted this alternative approach, using the head-turn preference procedure to examine the relative weight of phonetic and intensity changes on British English-learning 5-month-olds' word form recognition. Rather than presenting infants with alterations of their own name, each infant heard a variation of the same word. Restricting the stimulus set to one single word allowed for a controlled manipulation of both phonetic and intensity changes on the initial consonant and vowel. Specifically, the preference for the word *mummy*, (presumably recognised at this age (Tincoff & Jusczyk, 1999), was measured against either a phonetically or acoustically altered variation. Phonetic alterations consisted of a one-feature continuant change to the initial consonant (*nummy*), a one-feature stop change to the initial consonant (*bummy*), or a one-feature change to the initial vowel (*memmy*). Intensity alterations consisted of a 5dB energy increase that was isolated to either the initial consonant (***M****ummy*) or the initial vowel (*m****U****mmy*). This energy increase was selected based upon the results in Delle Luche et al. (2017), which found an average difference of intensity between the correct pronunciations and onset consonant mispronunciations of infants' own names of .10dB, with a 95% confidence interval (CI) of -2.8 to 2.6. This broadly corresponded to the 5dB intensity change applied to the stimuli in the present experiment.

Based on previous findings in British English-learning 5-month-olds (Delle Luche et al., 2017), it could be expected that infants here would not discriminate between the correct pronunciation of a familiar word form and any phonetic change. Alternatively, given that the present study uses more controlled stimuli, infants may demonstrate a vowel bias in word form recognition that is found in their French-learning peers (Bouchon et al., 2015; Von Holzen & Nazzi, 2020). However, if British English-learning 5-month-olds' processing of consonants and vowels is driven by energy cues rather than phonetic cues (Delle Luche et al., 2017), then infants should prefer listening to a correct pronunciation when the intensity of the initial consonant or vowel has been increased. This effect may also be more pronounced for intensity changes made to the initial vowel, due to the greater acoustic saliency of vowels.

3.1.2 Method

Five groups of British English-learning 5-month-olds were tested on how phonetic or acoustic changes impacted their word form recognition. Using the head-turn preference task, infants heard tokens of the familiar word *mummy* alongside either a phonetic or intensity change on the initial consonant or the initial vowel. The procedure was comparable to previous studies investigating how phonetic changes impact word form recognition at 5 months (Bouchon et al., 2015; Delle Luche et al., 2017).

Participants

A total of 120 healthy 5-month-old infants were tested. All infants were British English-learning monolinguals, raised in the South West of England. See Table 2 for further information on the infants included in the experiment. Sample size for this experiment, as well as for all studies detailed in this thesis, was based upon previous research conducted on similar issues and at similar ages (Bouchon et al., 2015; Delle Luche et al., 2017;

Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020). The data of an additional 56 infants were excluded from the analysis due to crying or being distracted (*n* = 37), having two consecutive trials with looking times below 2 seconds, or having three or more of such trials in total (*n* = 9), parental interference (*n* = 3), or being an outlier (i.e., the mean orientation times were 2 SDs below or above the group mean; *n* = 7).

Table 2: *Participant information* (*age and gender) and details of experimental conditions in Experiment 1.*

Conditions (all n = 24)	Change	Phonemic	Age in Days	Female/Male
Phonetic: consonant (continuant-continuant)	**n**ummy	**n**ʌmɪ	155 (12)	9/15
Phonetic: consonant (continuant-stop)	**b**ummy	**b**ʌmi	155 (11)	10/14
Phonetic: vowel	m**e**mmy	m**ɛ**mi	159 (12)	14/10
Intensity: consonant	**M**ummy	**m**ʌmi	152 (10)	13/11
Intensity: vowel	m**U**mmy	m**ʌ**mi	151 (10)	12/12

Note. Standard deviations are reported in parentheses.

Materials

Each infant heard lists of tokens consisting of the correct pronunciations of the word *mummy* (/mʌmi/), with either a phonetic alteration or an intensity alteration. Phonetic alterations initially consisted of a one-feature continuant change to the word's initial consonant (*nummy*; /nʌmi/) and a one-feature change to the word's initial vowel (*memmy*; /mɛmi/). A further phonetic change was later added based upon the results of these conditions, consisting of a one-feature stop change to the initial consonant (*bummy*; /bʌmi/). Intensity alterations consisted of an increase in energy of 5dB on either the initial consonant or the initial vowel on each correctly produced token (normalised at 65dB). To achieve the increases in intensity, the speech analysis software Praat (Boersma &

Weenink, 2010) was used to segment and isolate the initial consonant (i.e., ***M**ummy*; /**m**ʌmi/) and the initial vowel (i.e., *m**U**mmy*; /m**ʌ**mi/) of each correct pronunciation. Following this, an intensity tier was created in Praat, with intensity points then added around either the isolated initial consonant or the isolated initial vowel to increase its intensity by 5dB. See Figure 3 for an example.

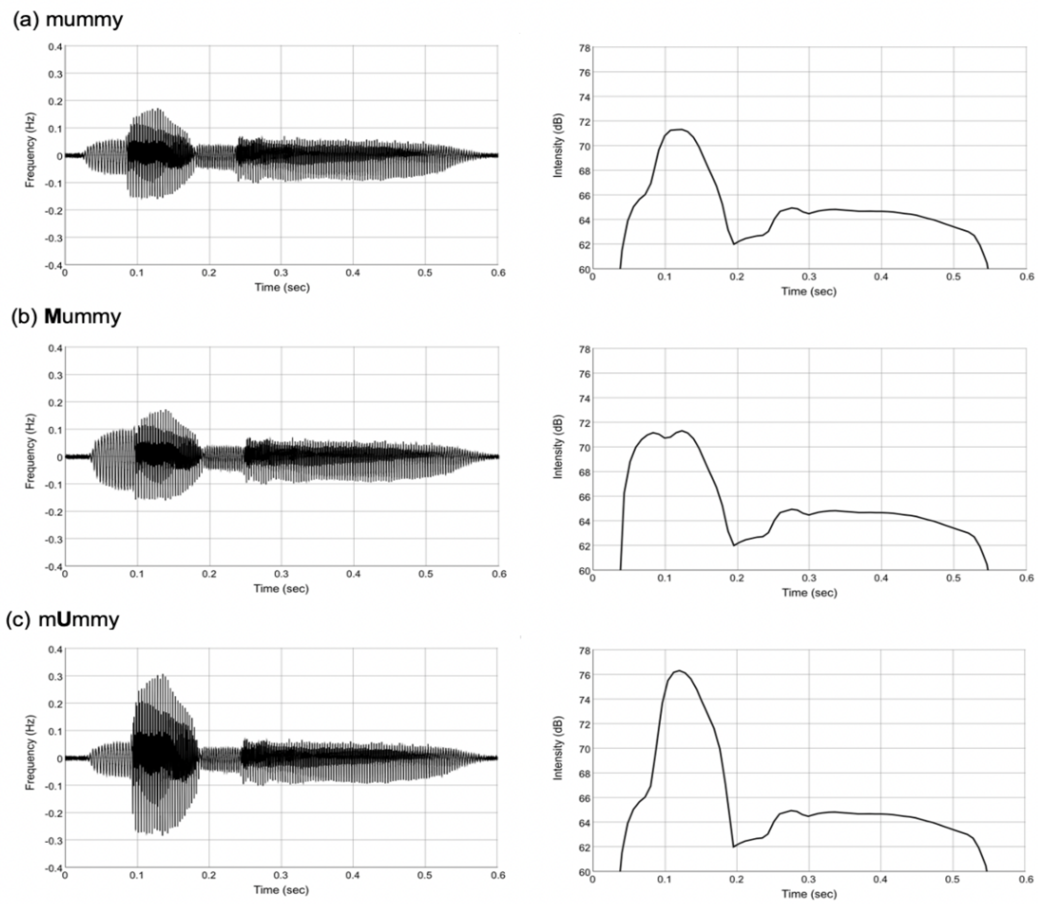


Figure 3: *Example time waveforms (left) and intensity contours (right) for a 'mummy' token used in the intensity conditions for (a) the unaltered token, (b) 5dB added to the onset consonant (/m/), and (c) 5dB added to the onset vowel (/ʌ/).*

The stimuli were recorded using a Zoom H4N Pro digital recorder in a soundproof booth across two different sessions by a native British English female speaker. All tokens were spoken in an infant directed voice. From the approximately 30 tokens of each word produced, 15 were selected, based primarily on an equal variety of intonation patterns. In the first session, the tokens *mummy*, *nummy*, and *memmy* were recorded. Table 3a shows the acoustic characteristics of these stimuli. Acoustic analysis revealed no statistical differences in mean, minimum, or maximum fundamental frequency between *mummy* and *nummy* and *mummy* and *memmy*. However, *mummy* tokens were longer in duration compared to *memmy*, In the second session, the word *bummy* was recorded, as well as an additional set of *mummy* tokens for this condition. Table 3b displays the acoustic characteristics of these tokens. Acoustic analysis here found no statistical differences in duration, minimum, or maximum fundamental frequency between the two words. However, these tokens of *mummy* had a significantly higher pitch compared to *bummy*. Pseudo-randomised lists were created for each condition, with an inter-stimulus interval (ISI) of 1 second between tokens. The lists were 24 seconds in length in each condition.

Table 3: *Acoustic features of stimuli presented in Experiment 1.*

(*a)*

	Condition			t-value and significance level (two-sided)
	mummy	nummy	memmy	
Duration (ms)	551 (106)	547 (91)	472 (39)	$t(28) = .10$, $p = .92$, $t(28) = 2.71$, $p = .01$
F0 mean (Hz)	307 (40)	314 (59)	302 (38)	$t(28) = -.43$, $p = .67$, $t(28) = .30$, $p = .76$
F0 min (Hz)	253 (55)	265 (67)	255 (53)	$t(28) = -.51$, $p = .62$, $t(28) = -.09$, $p = .93$
F0 max (Hz)	383 (42)	394 (74)	370 (44)	$t(28) = -.48$, $p = .63$, $t(28) = .85$, $p = .40$

Note. Standard deviations are reported in parentheses. The first t-test on each line corresponds to the comparison between *mummy* and *nummy*, and the second to the comparison between *mummy* and *memmy*.

(*b*)

	Condition		t-value and significance
	mummy	bummy	level (two-sided)
Duration (ms)	479 (72)	461 (46)	$t(28) = .78$, $p = .44$
F0 mean (Hz)	263 (30)	240 (13)	$t(28) = 2.71$, $p = .01$
F0 min (Hz)	216 (29)	203 (16)	$t(28) = 1.54$, $p = .14$
F0 max (Hz)	322 (50)	292 (43)	$t(28) = 1.76$, $p = .09$

Note. Standard deviations are reported in parentheses. T-test corresponds to the comparison between *mummy* and *bummy*.

Procedure

Following informed consent being provided by the parent, infants were seated on their parent's lap in a sound-attenuated, darkened booth (see Appendix A for a diagram of the experimental setup). A green light at the infants' eye level, with a video camera used to monitor the infant, was attached to a central panel in the front of the booth. Red lights, with loudspeakers below them, were located on panels either side of the booth. Parents were instructed to wear headphones playing a mix of speech and music to mask the auditory stimuli and prevent any inadvertent influence on the infants' looking behaviour. The experimenter sat outside the booth at the computer and video monitor used to control the stimulus presentation and record the infant looking times. Experimenters were unaware of the sound being played in the booth.

At the beginning of each trial, the light directly in front of the infant flashed green until the experimenter deemed the infant to be looking at it, at which point the red light on either the left or right would begin to flash. When the experimenter judged the infant to be attending to the side light, the sound file was played until its conclusion or until the infant failed to maintain their gaze towards the corresponding flashing light for two consecutive seconds. The cycle would then begin again. In trials where the infants stopped attending to the light for less than 2 seconds before turning back again, the sound file continued to play although the time spent looking away was automatically subtracted from the total looking time. Therefore, each trial had a maximum looking time of the entire sound file.

The session consisted of two practice trials (passages of classical music), presented on each side, to allow infants to become familiar with the procedure. This was followed by a test phase of eight trials, organised into two blocks. Each block contained two lists of the correctly pronounced word *mummy* and an alteration. The order the lists were presented in each block was randomised. A between-participant design was used in which each child was presented with one of the 5 conditions (continuant consonant phonetic change, stop consonant phonetic change, vowel phonetic change, consonant energy change, vowel energy change).

3.1.3 Results and discussion

Trials with looking times of under 2 seconds in duration were removed before statistical analysis. This led to the exclusion of 30 of the 960 test trials (3.13%). All infants included in the analysis contributed at least 6 out of 8 useable trials. Mean raw looking times to the correct pronunciation and altered pronunciation in each condition are presented in Figure 4.

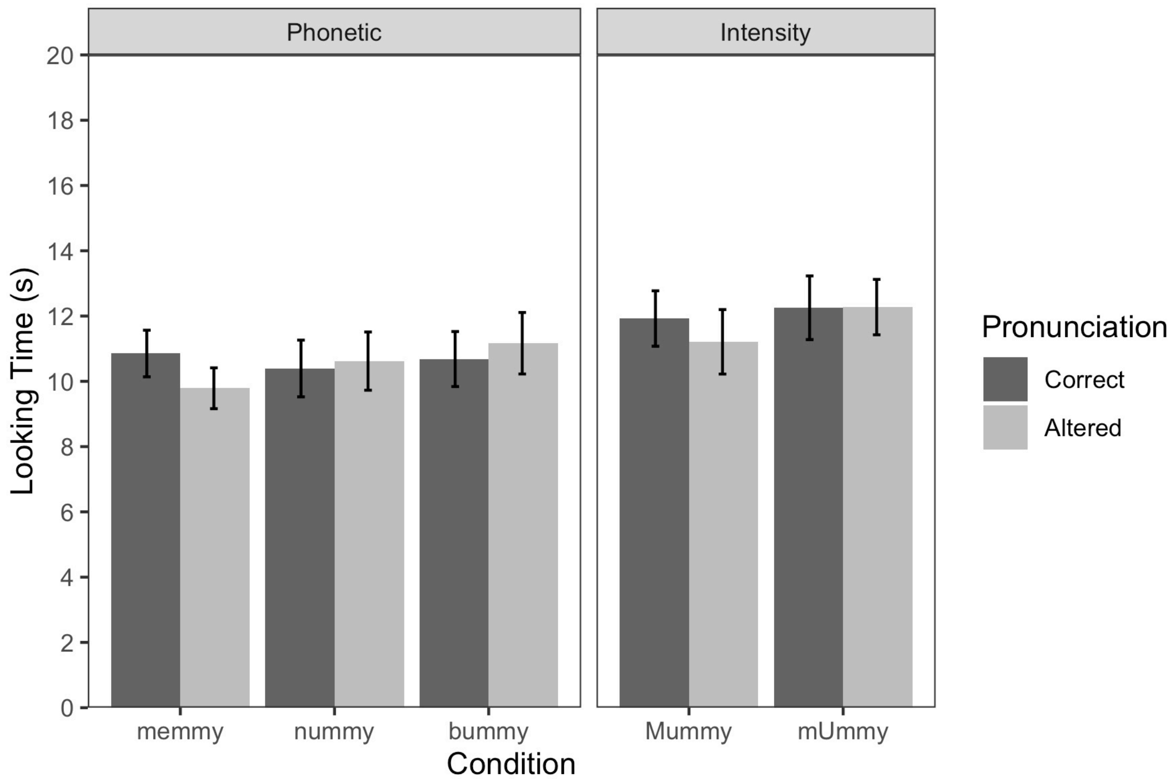


Figure 4: *Mean looking time* (*s) in Experiment 1 toward the correct pronunciations* (*dark grey bar) versus the altered pronunciations* (*light grey bar) of the word mummy, in the phonetic change condition* (*left) and intensity change condition* (*right). Brackets represent ±1 standard error.*

Looking times were not normally distributed, as revealed by a significant Shapiro-Wilk test (W = .97, p = .01 in Correct; W = .96, p < .001 in Altered). Therefore, infants' raw looking times were log-transformed as per the recommendations detailed in Csibra et al. (2016).

Phonetic conditions

Focussing first on the continuant consonant change of *nummy* and the vowel change of *memmy*, a three-way mixed analysis of variance (ANOVA) was conducted on infants' log-transformed looking times, with the within-subject factors of Pronunciation (correct versus altered) and Block (first half versus second half), and the between-subject factor of Condition (*nummy* versus *memmy*). This found that neither the effects of Pronunciation, $F(1, 46) = 1.15$, $p = .29$, $\eta_G^2 = .003$, or Condition, $F(1, 46) = .12$, $p = .92$, $\eta_G^2 = .001$, nor the Pronunciation x Condition interaction, $F(1, 46) = .47$, $p = .50$, $\eta_G^2 = .001$, were significant. There was a significant effect of Block, with infants looking longer during the first half of the experiment, $F(1, 46) = 42.26$, $p < .001$, $\eta_G^2 = .13$. However, there was no significant Block x Pronunciation, $F(1, 46) = .16$, $p = .69$, $\eta_G^2 < .001$, Block x Condition, $F(1, 46) = .002$, $p = .97$, $\eta_G^2 < .001$, or Block x Condition x Pronunciation interaction, $F(1, 46) = 1.18$, $p = .28$, $\eta_G^2 = .004$, indicating that infants' preferences remained the same in both the first and second half of the experiment.

Given that infants' preferences were not significantly impacted by Block, the statistical analysis on infants' looking times was rerun excluding block from the ANOVA. This found that neither the effects of Pronunciation, $F(1, 46) = .72$, $p = .40$, $\eta_G^2 = .003$., or Condition, $F(1, 46) = .003$, $p = .96$, $\eta_G^2 < .01$, nor the Pronunciation x Condition interaction, $F(1, 46) = 1.42$, $p = .24$, $\eta_G^2 = .01$, were significant.

Next, the stop consonant change of *bummy* versus the vowel change of *memmy* was analysed. A three-way mixed ANOVA on infants' looking times, with the within-subject factors of Pronunciation (correct versus altered) and Block (first half versus second half of the experiment) and the between-subject factor of Condition (*bummy* versus *memmy*) again found that the effects of Pronunciation, $F(1, 46) = .12$, $p = .73$, $\eta_G^2 < .001$, Condition, $F(1, 46) = .03$, $p = .87$, $\eta_G^2 < .001$, and the Pronunciation x Condition interaction, $F(1, 46) = 2.03$, $p = .16$, $\eta_G^2 = .01$, were not significant. There was a significant effect of Block, with looking times longer during the first half of the experiment, $F(1, 46) = 45.97$, $p < .001$,

η_G^2 = .11. However, there was no significant Block x Pronunciation, $F(1, 46) = .56$, p = .46, η_G^2 = .002, Block x Condition, $F(1, 46)$ = ..95, p = .33, η_G^2 = .003, or Block x Condition x Pronunciation interaction, $F(1, 46) = .66$, $p = .42$, η_G^2 = .002, suggesting that infants' behaviour remained the same in both blocks of the experiment.

Again, as the effect of Block did not influence infants' behaviour, statistical analysis on infants' looking times was rerun excluding Block from the ANOVA. This found that neither the effects of Pronunciation, $F(1, 46) = .38$, $p = .54$, η_G^2 = .001, or Condition, $F(1, 46) = .07$, $p = .79$, η_G^2 = .001, nor the Pronunciation x Condition interaction, $F(1, 46)$ = 1.73, $p = .20$, η_G^2 = .01, were significant.

The reliability of this null finding (no effect of condition) was further assessed by performing a paired Bayesian *t*-test, examining preferences for the correct pronunciation of mummy versus each phonetic change. A Bayes factor over 3 provides support for the strength of the alternative hypothesis (i.e., the presence of an effect) whereas a Bayes factor below 1/3 provides support for the strength of the null hypothesis (i.e., an absence of an effect). Any value that is between 1/3 and 3 is deemed to be inconclusive, providing no clear evidence for either the alternative or null hypothesis (Dienes, 2014; Jeffreys, 1961). To increase the objectivity of the analysis and allow a direct comparison between each experiment in the thesis, all Bayesian paired t-tests were calculated with the default Cauchy prior width of .707 in JASP (JASP Team, 2022; Version 0.16.3). Looking times for *nummy* versus *mummy* had a Bayes factor of BF = .22 ($t(23) = -.22$, $p = .83$), which provides evidence for the strength of the null hypothesis. Similarly, the comparison between *bummy* versus *mummy* revealed a Bayes factor of BF = .24 ($t(23) = -.44$, p = .67), again suggesting evidence for the absence of any effect. Finally, looking times for *memmy* versus *mummy* had a Bayes factor of BF = .66 (t(23) = 1.61, p = .12), which provides inconclusive evidence for either the alternative or null hypothesis.

Intensity conditions

Finally, the impact of increased intensity on the consonant (***M****ummy*) and the vowel (*m****U****mmy*) on infants' preferences was analysed. A three-way mixed ANOVA on looking times, with the within-subject factors of Pronunciation (correct versus altered) and Block (first half versus second half of the experiment), and the between-subject factor of Condition (***M****ummy* versus *m****U****mmy*) found no significant effect of Pronunciation, F(1, 46) = .61 p = .44, η_G^2 = .002. or Condition, F(1, 46) = .59, p = .45, η_G^2 = .01, nor was a significant Pronunciation x Condition interaction found, F(1, 46) = .84, p = .37, η_G^2 = .003. There was a significant effect of Block, with looking times longer during the first half of the experiment, F(1, 46) = 21.69, p < .001, η_G^2 = .05. However, there was no significant Block x Pronunciation, F(1, 46) = 1.98, p = .16, η_G^2 = .01, Block x Condition, F(1, 46) = 2.85, p = .10, η_G^2 = . 01, or Block x Condition x Pronunciation interactions, F(1, 46) = 2.04, p = .16, η_G^2 = .01, indicating that looking behaviours were consistent across the two blocks.

The ANOVA was rerun without the effect of Block, finding that neither Pronunciation, F(1, 46) = .53, p = .47, η_G^2 = .002., Condition, F(1, 46) = .49, p = .49, η_G^2 = .01, nor the Pronunciation x Condition interaction, F(1, 46) = 1.34, p = .25, η_G^2 = .01, were significant.

As with the phonetic conditions, Bayes factors were again calculated to examine the reliability of the null findings. The comparison between ***M****ummy* and *mummy*, found a Bayes Factor of BF = .42 (t(23) = 1.24, p = .23), providing inconclusive evidence for either the alternative or null hypothesis. Looking times between *m**U**mmy* and *mummy* revealed a Bayes factor of BF = .23 ($t(23) = -.33$, $p = .74$), providing evidence for the strength of the null hypothesis.

The results of Experiment 1 provide further evidence that, at 5 months, British English-learning infants' recognition of familiar word forms does not yet appear to be phonetically specified. Rather than presenting each infant with phonetic alterations of their own name, as in previous research (Bouchon et al., 2015; Delle Luche et al., 2017; Von Holzen & Nazzi, 2020), infants heard phonetic changes made to the familiar word *mummy.* However, infants did not appear to detect any one-feature changes made to this different word form, attending equally to both the correct pronunciation and the phonetic change. This was irrespective of whether the change was made to the initial consonant, for both stop and continuant alterations, or to the initial vowel. Note that infants' overall listening times to the correct and altered pronunciations of the word *mummy* were comparable to infants' overall listening times to the correct and altered pronunciations of their own name reported in Delle Luche et al. (2017). This suggests that infants were showing a similar preference for both their own name and the word *mummy*. Experiment 1 demonstrates that British English-learning infants behave differently to their French-learning peers, who show a vowel bias at 5 months when one-feature phonetic changes are made to their own name (Bouchon et al., 2015; Von Holzen & Nazzi, 2020). Instead, the results replicated previous research in British English-learning 5-month-olds, which found that consonant and vowel changes did not impact own name recognition (Delle Luche et al., 2017). This adds further support to cross-linguistic variations existing in the development of the consonant bias.

Furthermore, acoustic changes made to the stimuli did not appear to impact infants' preferences. When presented with controlled energy alternations, made to either the initial consonant or the initial vowel of each token, alongside unaltered versions of the familiar word form, infants again listened equally to both types of stimuli. This finding is different to previous research in British English-learning 5-month-olds. Delle Luche et al. (2017) found that infants used energy information in the consonants in a word recognition task. This was demonstrated by a higher tendency for infants to differentiate a correct pronunciation of their name from a mispronunciation when their name began with a stop consonant rather than a continuant consonant. However, the current finding that infants failed to discriminate the energy change, when using highly controlled stimuli, leads to the conclusion that there is no robust evidence for an early sensitivity to energy differences in consonants in British English-learning 5-month-olds. In contrast, there is instead strong evidence that infants learning this language are not able to process any kind of changes in consonants in word sequences at 5 months, whether it is a phonetic change or an acoustic change.

3.2 Experiment 2: Phonetic specificity of British English-learning 11-month-old infants' word form recognition

3.2.1 Introduction

The findings outlined in Experiment 1 suggest that, at 5 months, consonants and vowels matter equally to British English-learning infants' word form recognition. The aim of Experiment 2 was to examine the development of the consonant bias in slightly older

British English-learning infants, focussing on how 11-month-olds process consonant and vowel changes in lists of familiar words. It also explored whether lexical factors impacted infants' preferences. By 11 months, infants have presumably acquired their native vowel categories (e.g., Polka & Werker, 1994) and are becoming attuned to their native consonants (e.g., Werker & Tees, 1984). They have also acquired an average receptive vocabulary of approximately 135 words (Mayor & Plunkett, 2011). Furthermore, infants at 11 months show a preference for listening to familiar over novel words forms (e.g., Hallé & de Boysson-Bardies, 1994, 1996; Poltrock & Nazzi, 2015; Swingley, 2005; Vihman et al., 2004). Consequently, this age provides an ideal moment to investigate the role of consonants in lexical processing of familiar words and its potential link to vocabulary size.

The starting point for these experiments is the paradigm developed by Hallé and de Boysson-Bardies (1996) in French-learning infants, and then extended to English (Vihman et al., 2004) and Dutch (Swingley, 2005). Using a head-turn procedure, the authors examined the impact of mispronunciations of known words on infants' word form recognition. In 11-month-old English-learning infants, having first established that infants preferred listening to lists of disyllabic familiar word forms (e.g., *bubbles*) over disyllabic unfamiliar words, it was found that altering the initial consonants (e.g., *mubbles*), of the familiar words impeded word form recognition. Changing the onset of the familiar word's second syllable (e.g., *bummles*) impacted word recognition during the first half of the experiment, but not during the second half. This suggests that infants took longer to detect medial consonant mispronunciations. Replacing either the initial or final consonant in monosyllabic familiar words with a phonetically close consonant also resulted in the familiar word preference disappearing in Dutch-learning 11-month-olds (Swingley, 2005).

In French-learning 11-month-olds, Hallé and de Boysson-Bardies (1996) reported that only removing the onset consonant of a familiar disyllabic word resulted in a disappearance of the familiarity preference. However, a reanalysis of the data by Vihman et al. (2004) found that infants also recognised onset consonant mispronunciations but only during the second half of the trials and not the first. Such results suggest that infants' recognition of familiar words is at least partially disrupted by consonant changes.

Subsequent research using the same paradigm has demonstrated that word form recognition in French-learning 11-month-olds is impacted by consonant mispronunciations but not vowel mispronunciations. Poltrock and Nazzi (2015), having first established longer listening times to familiar over novel words, found that French-learning infants preferred to listen to vowel alterations over consonant alterations of the same familiar words at 11 months. This result was irrespective of the infants' overall lexicon size. Due to infants' preferences for familiar words at this age, this finding suggests a reliance on consonant information over vowel information in word form recognition. More recently, Von Holzen and Nazzi (2020) found that both 5- and 8-month-old French-learning infants displayed a vowel bias when recognising their own name, but that 11-month-olds displayed a consonant bias. This preference was also unrelated to infants' reported vocabulary size, as well as the proportion of unique consonant and vowel sequences that they knew. Together, these two studies demonstrate that a consonant bias appears to be present in French-learning infants in word form recognition by the age of 11 months. Furthermore, such findings suggest that lexical factors do not impact the reliance of consonants over vowels when performing such a task.

Nevertheless, research has yet to examine whether British English-learning 11-month-olds' word form recognition is impacted by consonant and vowel mispronunciations. In an IPL study, Mani and Plunkett (2007) found a potential consonant bias in word recognition at 15 months, when infants visually matched vowel but not consonant mispronunciations to one of two images. In contrast, 12-, 18-, and 24-month-old infants' word recognition was impacted by both phoneme changes on the same task (Mani & Plunkett, 2007, 2010). Combined with the finding that British English-learning infants do not display a consonant bias in word learning until 30 months (Floccia et al., 2014; Nazzi et al., 2009), this indicates that the observation of a consonant bias at 15 months in Mani and Plunkett (2007) could have been a statistical outlier. Based on this it could be predicted that 11-month-olds in the present study would show an equal preference for consonant and vowel changes when recognising familiar word forms. However, the word recognition tasks used in Mani and Plunkett (2007, 2010) required infants to demonstrate a meaningful relationship between a word form and its visual referent. In contrast, the word form recognition task used here only requires infants to identify the sound form of a familiar word. Thus, presenting infants with this simpler task may increase the opportunity of observing a consonant bias.

Using the head-turn preference paradigm, Experiment 2a aimed to replicate the finding that infants prefer listening to lists of familiar over unfamiliar disyllabic word forms, as has been previously demonstrated in the literature (e.g., Hallé & de Boysson-Bardies, 1994; Vihman et al., 2004). Having established this familiarity preference, Experiment 2b followed on from the findings in French-learning infants (Poltrock & Nazzi, 2015) by examining infants' preferences for onset consonant mispronunciations versus vowel mispronunciations of the familiar word forms presented in Experiment 2a. If British English-learning 11-month-olds behave in the same way as their French-learning counterparts, then a preference for lists of word forms with a vowel mispronunciation should emerge. However, if British English-learning infants do not show a consonant bias, then no preference for either consonant or vowel changes to the familiar word forms will be found. Experiment 2c also explored whether infants prefer listening to vowel mispronunciations compared to medial consonant mispronunciations.

Across all three experiments, infants' preferences were also measured in relation to their vocabulary size as estimated by parental reports, to examine the link between the acquisition of a consonant bias and lexical development. To refine these analyses, infants' vocabulary knowledge was also used to estimate the number of unique consonant tiers and vowel tiers infants had learned at this stage to examine if their knowledge of the structure of their native language predicted their preference for lists of items. This calculation followed the same logic as Keidel et al. (2007), who computed consonant and vowel tiers in adult lexicons, finding that words were around 40% more likely to be recognised based solely on their consonant tiers.

3.2.2 Experiment 2a: Familiar Word Forms versus Pseudowords

3.2.2.1 Introduction

To first establish a preference for familiar word forms, Experiment 2a used the head-turn preference task to measure 11-month-old British English-learning infants' preferences for lists of familiar words in comparison to lists of pseudowords. Due to previous research showing that infants at 11 months demonstrate a preference for familiar word forms over pseudowords (including for British English, Vihman et al., 2004), it was expected that infants would also show such a familiarity effect in the present experiment.

3.2.2.2 Method

Participants

A total of 24 healthy British English-learning monolingual 11-month-old infants were successfully tested (mean age = 10 months; 22 days, range = 10;8 days to 11;24, 12 females, 12 males). The data of 17 additional infants were excluded due to non-completion of the 12 trials due to fussiness (*n* = 3), having two consecutive trials with looking times below 2 seconds, or having three or more of such trials in total (*n* = 6), being an outlier (difference score below or above two standard deviations from the group mean; *n* = 4), and technical problems (*n* = 4). As a measure of infants' lexical development, parents were asked to complete the 100-word Oxford Short Form Communicative Developmental Inventory (CDI, Floccia et al., 2018; Hamilton et al., 2000), as well as a checklist of the 10 test words presented in the study. All infants were born and raised in the South West of England.

Materials

The lists of familiar words and pseudowords used are presented in Table 4. The ten disyllabic familiar words were chosen using the Oxford CDI (Hamilton et al., 2000). The selected words were comprehended by 38% (ranging from 14% to 70%) of British

English-learning 11-month-olds. This was comparable to the 30% comprehension of familiar words reported by Poltrock and Nazzi (2015), and 33% for Vihman et al. (2004). The rationale is that the words presented have a degree of familiarity for infants, sufficient to elicit word form recognition.

The nonwords were created by changing the first three phonemes of each familiar word (initial consonant, first vowel and middle consonant). For example. in *mummy* the initial consonant was changed to an /n/ to form *nummy*, and its first vowel to an /ɛ/ to form *nemmy*. The second consonant (e.g., the second /m/ of *mummy*) was also changed to avoid too much of an overlap between familiar words and pseudowords (in this example, /m/ was changed into /b/, so that the resulting pseudoword was *nebby* (/nɛbi/). No resulting nonword can be a real word in the dialect spoken in the South West of England where the experiments took place.

Table 4: *Familiar words and pseudowords, along with their phonemic transcriptions, presented to infants in Experiment 2a. Pseudowords were formed by altering the first three phonemes (initial consonant, first vowel and middle consonant) of each familiar word.*

Word	Phonemic Transcription of Word	Pseudoword	Phonemic Transcription of Pseudoword
baby	beɪbɪ	pyppy	paɪpɪ
bottle	bɒtl	puckle	pʌl
bunny	bʌni	pammy	pæmi
button	bʌtn	meddon	medn
cuddle	kʌdl	gannel	ganl
daddy	dædi	tenny	tenɪ
mummy	mʌmi	nebby	nebɪ
nappy	næpi	dubby	dʌbi
tickle	tɪkl	keggle	kegl
water	wɔ:tə	mirper	mɜ:pa

The stimuli were recorded in an infant-directed voice by a British English native speaker (from the South West of England) using a Zoom H4N Pro digital recorder in a soundproof booth. One token for each word and nonword was selected. The amplitude of all tokens was normalised to 70dB using Praat. Acoustic analysis of the recorded tokens found no statistical differences in duration, or mean, minimum, or maximum fundamental frequency between the words and nonwords (see Table 5). Pseudorandomised lists were created for both the words and nonwords. Lists were made up of two blocks, with each token appearing once in each block, resulting in lists containing a total of 20 tokens. An ISI of 600ms was used between tokens. The position each token was presented in each list was evenly distributed both within and across each list. The lists were 21.24 seconds in length in both conditions (words versus pseudowords).

Table 5: *Acoustic features of stimuli presented in Experiment 2a.*

	Condition		t-value and significance level (two-sided)
	Words	Pseudowords	
Duration (ms)	462 (33)	462 (53)	$t(18) = .00$, $p = .99$
F0 mean (Hz)	351 (50)	344 (36)	$t(18) = .39$, $p = .70$
F0 min (Hz)	262 (83)	261 (64)	$t(18) = .03$, $p = .98$
F0 max (Hz)	451 (32)	446 (20)	$t(18) = .45$, $p = .66$

Note. Standard deviations are reported in parentheses. Each *t*-test corresponds to the comparison between words and pseudowords.

Calculating consonant and vowel tiers

The proportion of consonants and vowels in words known by each infant was evaluated using the same method as Von Holzen and Nazzi (2020). Infants' total word comprehension was determined by combining the parents' responses to the 100-word Oxford Short Form CDI and the 10 test words on the study checklist. This resulted in a potential 108 unique words (2 words appeared in both the CDI and the study checklist) that infants could be reported to comprehend. Phonetic transcriptions of the known words were then used to create consonant and vowel tier scores. For instance, if an infant understood the words daddy (/dædi/), mummy (/mʌmi/), and bunny (/bʌni/), then they would know three distinct consonant tiers, as all three words contain unique consonant sequences (/d.d/; /m.m/; /b.n/). However, they would only know two distinct vowel tiers, with mummy and bunny containing the same vowel sequence (/ʌ.i/), which is distinctive from the vowel sequence in daddy (/æ.i/). Consonant and vowel proportion scores were then calculated for each infant by dividing the number of unique consonant or vowel tiers they knew by their total word comprehension. These data were then used to correlate with looking times for the different types of word lists.

Procedure

The experiment used the same procedure and apparatus as detailed in Experiment 1. However, as in Poltrock and Nazzi (2015), Experiment 2a had a test phase consisting of 12 trials, six of each condition (familiar word lists and pseudoword lists). Trials were organised into two blocks. Each block contained three lists of familiar words and three lists of pseudowords. The order the lists were presented in each block was randomised.

3.2.2.3 Results and discussion

Prior to analysing the data, trials with looking times of under 2 seconds in duration were excluded. This led to the removal of 25 of the total 288 test trials (8.68%). All infants included in the analysis contributed at least 9 out of 12 useable trials. Each infant's mean raw looking times to the lists of familiar words (*M* = 7.74s, *SD* = 2.45) and pseudowords (*M* = 6.33, *SD* = 1.74) were calculated and are displayed in Figure 5. Shapiro-Wilk tests for normality revealed that infants' looking times to familiar words (W = .96, p = .63) and pseudowords (W = .97, p = .67) were normally distributed.

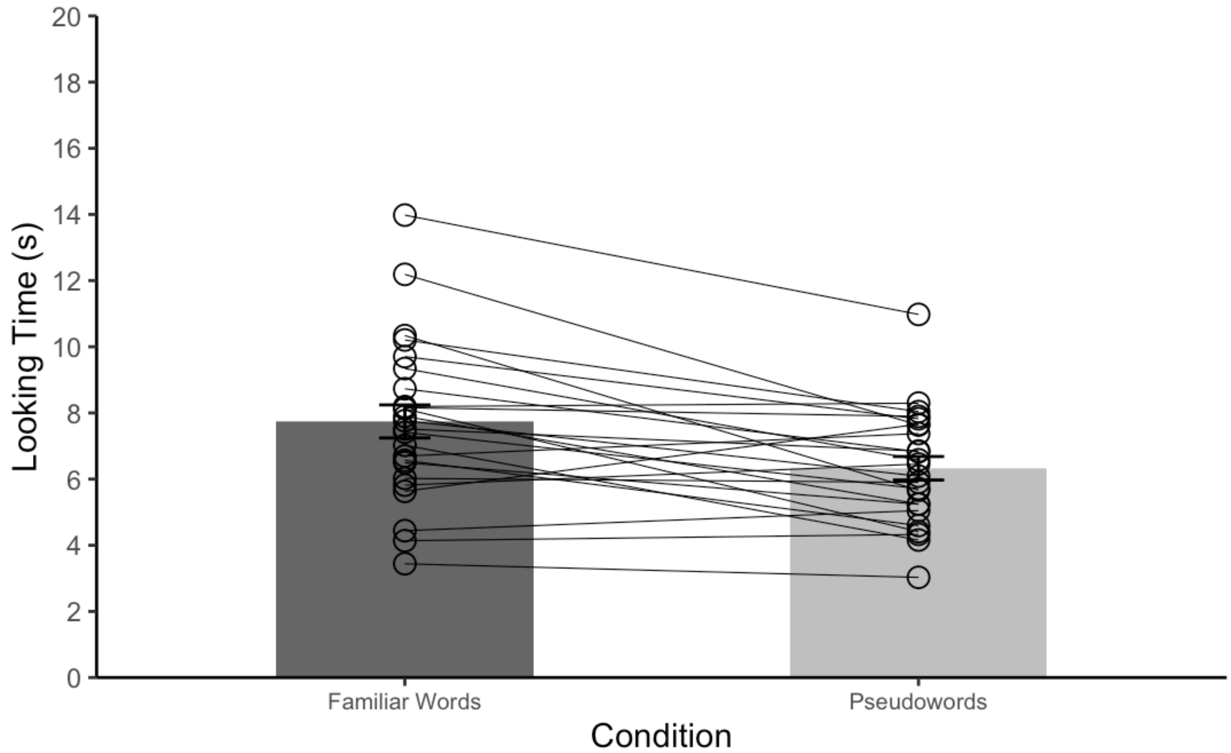


Figure 5: *Mean looking times* (*s*) *in each condition in Experiment 2a* (*familiar word forms versus pseudowords). Connected dots represent individual infants' looking times in the two experimental conditions. Error bars represent ±1 standard error.*

A repeated measures ANOVA on the orientation times with List (familiar words versus pseudowords) and Block (first half versus second half of the experiment) as within-participant factors found a main effect of List, $F(1, 23) = 15.83$, $p = .001$, $\eta_G^2 = .41$, with infants listening longer to the lists of familiar words (M = 7.74s, SD = 2.45) than pseudowords (M = 6.33, SD = 1.74). A significant effect of Block was also found, $F(1, 23) = 12.11$, $p = .002$, $\eta_G^2 = .35$, with infants listening longer during the first half of the study. However, there was no List x Block interaction, $F(1, 23) = .68$, $p = .42$, $\eta_G^2 = .03$.

The reliability of this finding was further examined with a paired Bayesian *t*-test, comparing the looking times for familiar words versus pseudowords. This found a Bayes factor of BF = 58.01 ($t(23) = 4.00$, $p < .001$), providing evidence for the alternative hypothesis.

Bayesian correlations were also conducted to examine the relationship between infants' listening preferences (calculated as the difference in mean orientation times to the familiar words and pseudowords) and lexical factors. All Bayesian correlations in the paper were calculated using the default stretched beta prior width of 1.00 in JASP (JASP Team, 2022). Infants had an average CDI comprehension of 7.71 (*SD* = 5.89) out of 100 words and knew an average of 4.58 (*SD* = 2.21) out of the 10 test words presented in the study. No correlation was found between infants' listening preferences and either CDI ($r(22) = .12$, $p = .57$, BF = .30) or word checklist scores ($r(22) = -.08$, $p = .70$, BF = .27), indicating evidence in favour of the null hypothesis. From the average of 11.00 words (*SD* = 6.84) that infants understood (from the 108 unique words found when combining the Oxford Short Form CDI and the study checklist), there was an average 10.54 (*SD* = 6.41) unique consonant tiers and 9.54 (*SD* = 5.18) unique vowel tiers. The proportion of unique consonant tiers out of known words (*M* = .97, *SD* = .05) was significantly higher than the proportion of vowel tiers (*M* = .90, *SD* = .10), $t(23) = 2.73$, $p = .01$. Neither proportion of consonant ($r(22) = -.09$, $p = .69$, BF = .27) nor vowel ($r(22) = .01$, $p = .98$, BF = .25) tiers out of known words was correlated with infants' listening preferences, again providing evidence in favour of the null hypothesis. Hence, this analysis fails to find correlations between the proportion of unique consonant or vowel tiers in the words known by each child and their listening preference for familiar words over pseudowords.

The findings of Experiment 2a demonstrate that British English-learning 11-month-olds listened longer to the list of familiar word forms over pseudowords. This preference was irrespective of the reported CDI, word checklist scores, or the proportion of vowels and consonants in known words. Therefore, infants were able to either comprehend or recognise a sufficient amount of the words presented to them to display an overall preference for familiar word forms over pseudowords. This result replicates previous word form recognition studies at this age (Hallé & de Boysson-Bardies, 1994, 1996; Poltrock & Nazzi, 2015; Swingley, 2005; Vihman et al., 2004), providing further support that, even at this early stage of development, infants can recognise the auditory form of familiar words in their environment.

3.2.3 Experiment 2b: Onset consonant changes versus vowel changes

3.2.3.1 Introduction

The results of Experiment 2a showed that British English-learning 11-month-old infants preferred listening to familiar word forms over pseudowords, irrespective of parental reports of their comprehension of such words. In Experiment 2b, British English-learning 11-month-olds' preference for either an onset consonant mispronunciation or vowel mispronunciation of the familiar word forms presented in Experiment 2a was examined. Given that most disyllabic words in English have a stress that is word initial (Cutler & Carter, 1987), as is the case for all the familiar words in the present stimulus set, it can be predicted that the initial phoneme mispronunciations would be particularly salient to the infants in the present study. This is supported by the finding that American English-learning 9-month-olds listen significantly longer to stress-initial over stress-final words (Jusczyk et al., 1993). Furthermore, based on Poltrock and Nazzi's (2015) finding that

French-learning 11-month-olds prefer listening to vowel mispronunciations over consonant mispronunciations of familiar word forms, if British English-learning 11-month-olds demonstrate a consonant bias, then it could be expected that they will show the same preference for vowel mispronunciations compared to consonant mispronunciations of the word forms presented in Experiment 2b. However, if they showed no bias at this age, then listening times would be equal to both vowel mispronunciations and consonant mispronunciations of such familiar words.

3.2.3.2 Method

Participants

A further sample of 24 healthy British English-learning monolingual 11-month-old infants were successfully tested (mean age = 11;1, range = 10;11 to 11;28, 8 females, 16 males). The data of four additional infants were excluded due to fussiness (n = 1) and having two consecutive trials with looking times below 2 seconds or having three or more of such trials in total (n = 3). Parents completed both the 100-word Oxford Short Form CDI (Floccia et al., 2018; Hamilton et al., 2000) and a checklist of the correct pronunciation of the 10 test words presented in the study. All infants were born and raised in the South West of England.

Materials

The familiar words from Experiment 2a were modified by one phonological feature, either on the first consonant or the first vowel. The types of changes were the same as those used in Experiment 2a to create pseudowords. That is, whereas the word *mummy* was changed to *nebby* in Experiment 2a by changing the first three phonemes, here only the first consonant (*nummy*) or the first vowel (*memmy*) was altered. As in Poltrock and Nazzi (2015), several types of phonological feature changes were presented in order to reflect consonant and vowel categories as a whole (see Table 6 for the list of consonant and vowel changes to the familiar words and type of phonological feature changes).

Table 6: *Onset consonant and vowel changes to familiar words presented in Experiment 2b.*

	Onset Consonant Change			Vowel Change		
Word	Change	Phonemic	Type	Change	Phonemic	Type
baby	paby	pəɪbɪ	place	byby	baɪbi	height
bottle	pottle	pɒtl	voice	buttle	bʌtl	backness
bunny	punny	pʌni	place	banny	bænɪ	backness
button	mutton	mʌtn	manner	betton	betn	backness
cuddle	guddle	gʌdl	voice	caddle	kədl	backness
daddy	taddy	tædɪ	voice	deddy	dedɪ	height
mummy	nummy	nʌmɪ	place	memmy	memɪ	backness
nappy	dappy	dæpɪ	manner	nuppy	nʌpɪ	backness
tickle	kickle	kɪkl	place	teckle	tekl	height
water	mawter	mɔːtə	manner	wirter	wɜːtə	backness

The tokens were recorded using the same speaker and recording arrangement as in Experiment 2a. The acoustic features of the stimuli are listed in Table 7. There were again no significant differences in duration or, mean, minimum, or maximum fundamental frequency in the sound files for each condition. The amplitude of each token was normalised to 70dB using Praat. The mispronunciations were placed into lists using the same procedure and ISI as Experiment 2a, with each list being 21.92 seconds in length in both conditions.

Table 7: *Acoustic features of stimuli presented in Experiment 2b.*

	Condition		t-value and significance level (two-sided)
	Onset Consonant Change	Vowel Change	
Duration (ms)	496 (50)	494 (52)	$t(18) = .10$, $p = .92$
F0 mean (Hz)	350 (57)	345 (52)	$t(18) = .20$, $p = .84$
F0 min (Hz)	275 (90)	267 (86)	$t(18) = .20$, $p = .84$
F0 max (Hz)	447 (24)	453 (18)	$t(18) = -.63$, $p = .54$

Note. Standard deviations are reported in parentheses. Each *t*-test corresponds to the comparison between onset consonant mispronunciations and vowel mispronunciations.

Procedure

The procedure and apparatus were identical to those used in Experiments 1 and 2a. The test phase consisted of 12 trials, six of each condition (onset consonant mispronunciation lists and vowel mispronunciation lists). Trials were again organised into two blocks. Each block contained three lists of onset consonant mispronunciation words and three lists of vowel mispronunciation words. The order of the lists within each block was randomised.

3.2.3.3 Results and discussion

As in Experiment 2a, trials with looking times of under 2 seconds in duration were excluded before analysing the data. This led to the exclusion of 20 of the 288 test trials

(6.94%). All infants included in the analysis contributed at least 9 out of 12 useable trials. Mean raw looking times to the lists of onset consonant change (M = 9.49s, *SD* = 2.93) and vowel change (M = 8.67s, *SD* = 3.03) to the familiar words were calculated and are displayed in Figure 6.

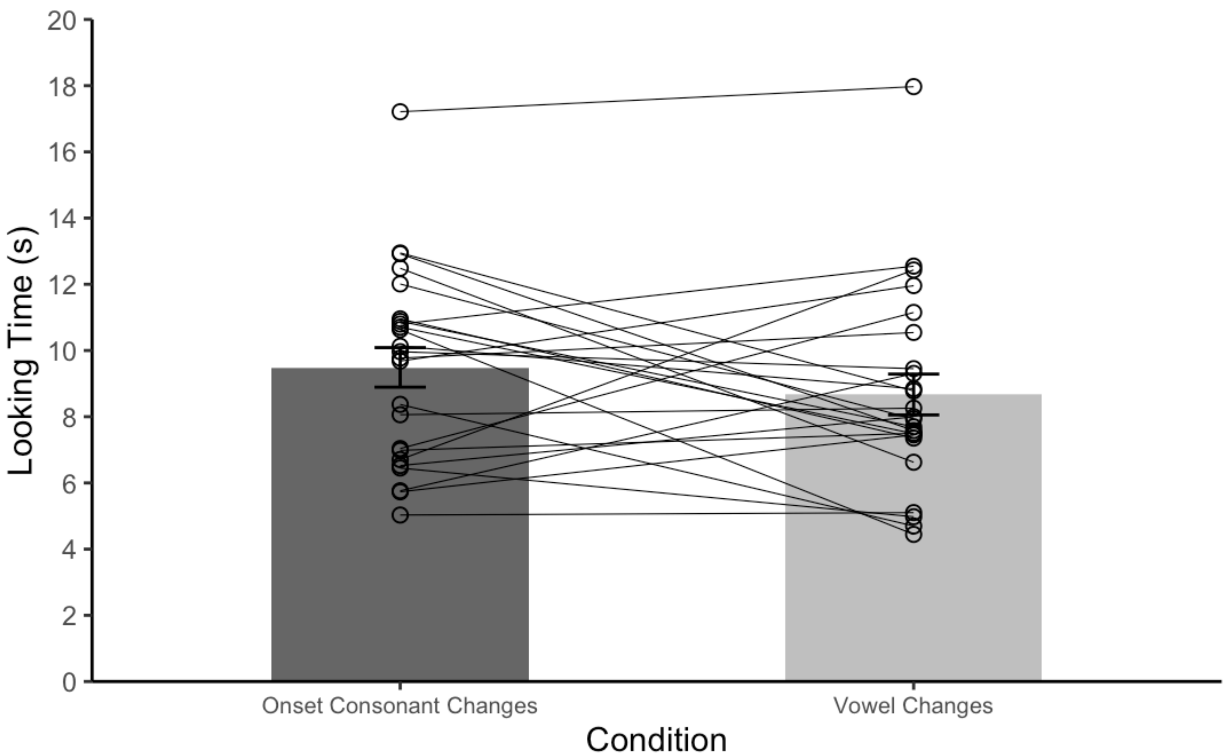


Figure 6: *Mean looking times (s) in Experiment 2b (onset consonant changes versus vowel changes). Connected dots represent individual infants' looking times in the two experimental conditions. Error bars represent ±1 standard error.*

Shapiro-Wilk tests for normality revealed that infants' looking times to onset consonant changes were normally distributed (W = .95, p = .22) but looking times to vowel changes were not normally distributed (W = .90, p = .03). Therefore, raw listening times were log-transformed prior to analysis (Csibra et al., 2016).

A repeated measures ANOVA on log-transformed orientation times with List (onset consonant change versus vowel change) and Block (first half versus second half of the experiment) as within-participant factors found a significant effect of Block, $F(1, 23) = 44.17$, $p < .001$, $\eta_G^2 = .18$, with infants again listening longer in the first half of the study. However, no main effect of List ($F(1, 23) = 1.51$, $p = .23$, $\eta_G^2 = .01$) or an interaction between List and Block ($F(1, 23) = .01$, $p = .92$, $\eta_G^2 < .001$) was found.

The reliability of the null finding (no effect of condition) was examined with a paired Bayesian *t*-test, comparing the log-transformed looking times for vowel mispronunciations versus consonant mispronunciations. This found a Bayes factor of BF = .43 ($t(23) = -1.25$, $p = .22$), providing inconclusive evidence for either the alternative or null hypothesis.

Infants had an average CDI comprehension of 6.67 words (*SD* = 5.02) out of 100 and knew an average of 5.08 (*SD* = 2.15) of the 10 words presented in the study. No correlation was found between infants' log-transformed looking preferences and CDI comprehension ($r(22) = -.07$, $p = .80$, BF = .26), providing evidence for the null hypothesis. However, inconclusive evidence for either the alternative or null hypothesis was found for the correlation between infants' listening preferences and word checklist score ($r(22) = .20$, $p = .34$, BF = .39). Infants knew an average of 10.42 words (*SD* = 6.56) out of the possible 108 unique words found when combining the Oxford Short Form CDI and the study checklist, from which there was an average 10.00 (*SD* = 6.20) consonant tiers and 9.00 (*SD* = 4.93) vowel tiers. The proportion of unique consonant tiers out of known words (*M* = .97, *SD* = .05) was found to be significantly different to the proportion of vowel tiers (*M* = .91, *SD* = .09), $t(23) = 2.81$, $p = .01$. However, neither proportion of consonant (*r*(22) = .02, *p* = .94, BF = .25) nor vowel tiers (*r*(22) = .04, *p* = .86, BF = .26) was correlated with infants' log-transformed looking preferences, suggesting evidence in favour of the null hypothesis.

Experiment 2a established that British English-learning 11-month-olds' word recognition was indexed by longer looking times towards familiar word forms. In this context, the ambiguity regarding infants' preferences for onset consonant mispronunciations, vowel mispronunciations, or neither, found here in Experiment 2 suggests that infants do not yet have a consonant bias in their word recognition. This is in contrast to Poltrock and Nazzi's (2015) finding that showed a consonant bias in word recognition in French-learning infants of the same age. The present results thus provide further evidence that infants' initial word recognition procedures may vary cross-linguistically, as previously found at later ages in word learning tasks (Floccia et al., 2014; Nazzi et al., 2016).

3.2.4 Experiment 2c: Medial consonant changes versus vowel changes

3.2.4.1 Introduction

In Experiment 2c, infants were again presented with consonant and vowel changes to the familiar words used in Experiment 2a. However, whilst the vowel changes remained the same as those used in Experiment 2b, the consonant changes here occurred on the medial consonant position of each word. These changes were made based on the findings of Vihman et al. (2004), who demonstrated that infants took longer to recognise familiar word forms when the medial consonant was altered, only showing a preference for such a mispronunciation over a pseudoword during the second half of trials. Given that no effect was observed in Experiment 2b when the consonant change was on the more salient initial position of the stressed syllable, it would seem unnecessary to carry out Experiment 2c. However, Experiment 2c was planned and executed in parallel to Experiment 2b, with no knowledge of the outcome of Experiment 2b at the time.

3.2.4.2 Method

Participants

A total of 24 healthy British English-learning monolingual 11-month-old infants were successfully tested (mean age = 11;1, range = 10;11 to 11;28 days, 8 females). The data of four additional infants were excluded due to fussiness (*n* = 1) and having at least three inattentions or two consecutive inattentions (*n* = 3). Parents completed both the 100-word Oxford Short Form CDI (Floccia et al., 2018; Hamilton et al., 2000) and a checklist of the words which were to be mispronounced in the experiment.

Materials

Tokens consisted of one phonological feature consonant or vowel changes to the familiar words presented in Experiment 2a. Again, several types of phonological feature changes were selected to reflect consonant and vowel categories as a whole (see Table 8 for the list of consonant and vowel changes to the familiar words and type of phonological feature change).

Table 8: *Medial consonant and vowel changes to familiar words presented in Experiment 2c.*

	Final Consonant Change			Initial Vowel Change		
Word	Change	Phonemic	Type	Change	Phonemic	Type
baby	bappy	beɪpɪ	place	byby	baɪbi	height
bottle	boddle	bɒdl	voice	buttle	bʌtl	backness
bunny	bummy	bʌmi	place	banny	bænɪ	backness
button	buddon	bʌdn	voice	betton	betn	backness
cuddle	cunnel	kʌnl	manner	caddle	kədl	backness
daddy	danny	dænɪ	manner	deddy	dedɪ	height
mummy	mubby	mʌbɪ	manner	memmy	memɪ	backness
nappy	nabby	næbi	place	nuppy	nʌpɪ	backness
tickle	tiggle	tɪgl	voice	teckle	tekl	height
water	wawper	wɔ:pə	place	wirter	wɜ:tə	backness

The tokens were again recorded using the same speaker and recording arrangements as previously discussed. The acoustic features for the stimuli are listed in Table 9. There were again no significant differences in duration, or mean, minimum, or maximum fundamental frequency in the sound files for each condition. All tokens were normalised for amplitude at 70dB using Praat. The mispronunciations were placed into lists using the same procedure and ISI as Experiment 2a, and 2b, with each list being 21.92 seconds in length in both conditions.

Table 9: *Acoustic features of stimuli presented in Experiment 2c.*

	Condition		t-value and significance level (two-sided)
	Medial Consonant Change	Vowel Change	
Duration (ms)	496 (58)	494 (52)	$t(18) = .10$, $p = .92$
F0 mean (Hz)	329 (47)	345 (52)	$t(18) = -.71$, $p = .48$
F0 min (Hz)	240 (56)	267 (86)	$t(18) = -.84$, $p = .41$
F0 max (Hz)	479 (42)	453 (18)	$t(18) = 1.77$, $p = .09$

Note. Standard deviations are reported in parentheses. Each *t*-test corresponds to the comparison between medial consonant mispronunciations and vowel mispronunciations.

Procedure

The procedure and apparatus were identical to those used in Experiments 1, 2a and 2b. The test phase consisted of 12 trials, six of each condition (medial consonant mispronunciations and vowel mispronunciations). Trials were again organised into two blocks. Each block contained three lists of medial consonant mispronunciation words and three lists of vowel mispronunciation words. The order the lists were presented in each block was again randomised.

3.2.4.3 Results and discussion

All trials with looking times of under 2 seconds in duration were excluded from the analysis of the data. This led to the exclusion of 14 of the 288 test trials (4.86%). All infants included in the analysis contributed at least 9 out of 12 useable trials. The raw mean listening times to the medial consonant change and vowel change lists were calculated for each infant (see Figure 7). Infants listened to the medial consonant change lists for an average of 6.54s (*SD* = 1.85) and the vowel change for 7.10s (*SD* = 1.99). Looking times were normally distributed, as revealed by a non-significant Shapiro-Wilk test (Medial Consonant W = .96, p = .36; Vowel W = .94, p = .17).

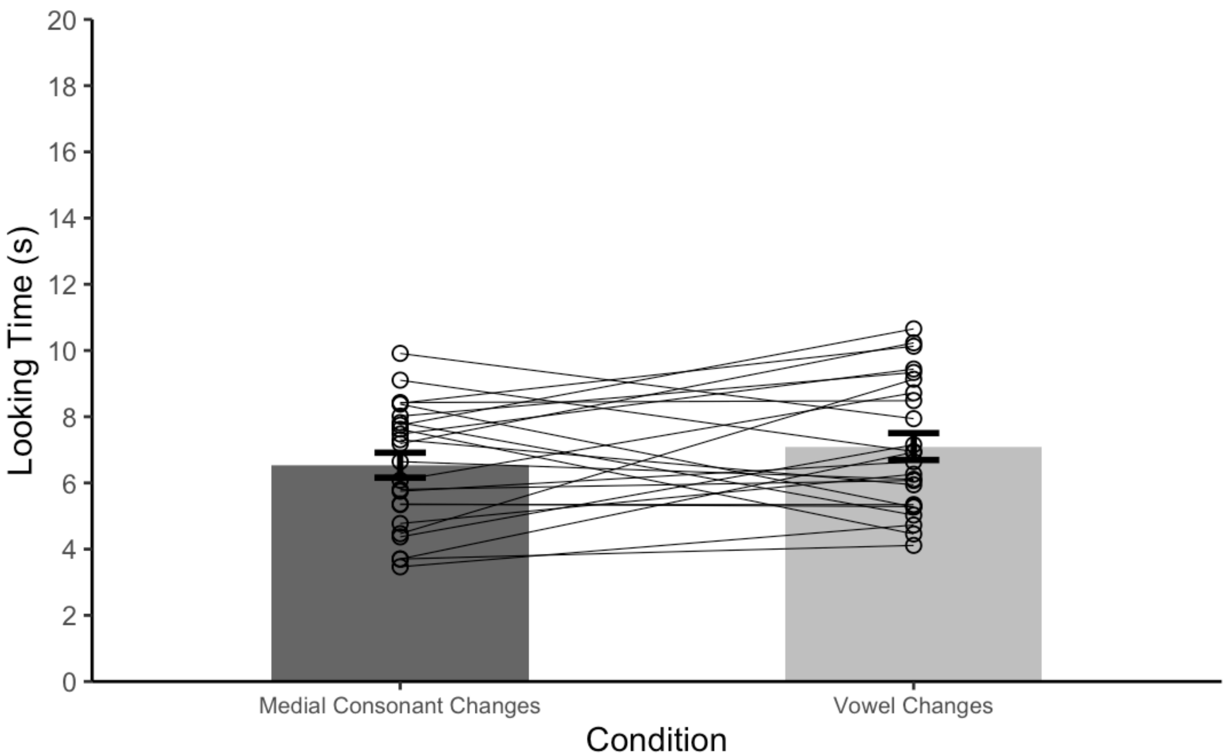


Figure 7: Mean looking times (s) in Experiment 2c (medial consonant changes versus vowel changes). Connected dots represent individual infants' looking times in the two experimental conditions. Error bars represent ±1 standard error.

A repeated measures ANOVA on the orientation times with List (medial consonant change versus vowel change) and Block (first half versus second half of the experiment) as within-participant factors revealed a significant effect of Block, $F(1, 23) = 31.61$, $p < .001$, $\eta_G^2 = .58$, with greater listening times found during the first half of the study. However, as in Experiment 2b, no main effect of List was found ($F(1, 23) = 1.71$, $p = .20$, $\eta_G^2 = .07$) and no interaction between List and Block ($F(1, 23) = .10$, $p = .76$, $\eta_G^2 = .004$) was found.

The reliability of the null finding (no effect of condition) was also examined with a paired Bayesian *t*-test, comparing the looking times for medial consonant mispronunciations and vowel mispronunciations. This found a Bayes factor of BF = .42 ($t(23) = -1.27$, $p = .22$), providing inconclusive evidence for either the alternative or null hypothesis.

Infants had an average CDI comprehension of 8.33 (*SD* = 5.80) words and knew an average of 5.79 (*SD* = 2.08) of the 10 words presented in the study. Again, no correlation was found between infants' listening time and either CDI ($r(22) = -.05$, $p = .82$, BF = .26) or word checklist ($r(22) = -.28$, $p = .19$, BF = .58) scores. In relation to consonant and vowel tiers, infants knew an average of 12.54 words (*SD* = 7.15), from which there was an average 11.92 (*SD* = 6.65) consonant tiers and 11.00 (*SD* = 5.30) vowel tiers. The proportion of unique consonant tiers (*M* = .96, *SD* = .05) was found to be significantly different to the proportion of vowel tiers (*M* = .92, *SD* = .10), $t(23) = 2.04$, $p = .05$. However, neither proportion of consonant ($r(22) = -.11$, $p = .62$, BF = .29) nor vowel tiers ($r(22) = -.15$, $p = .50$, BF = .32) was correlated with infants' looking time.

The results of Experiment 2c again appeared to show that British English-learning 11-month-olds' auditory recognition of known words was equally impacted by consonant and vowel changes. Infants did not demonstrate a preference for either a medial consonant mispronunciation or vowel mispronunciation to familiar words. Taken with the findings of Experiment 2b, this shows that the lack of a preference for vowel over consonant changes does not depend on the position of the consonant change at this age.

3.3 Experiment 3: Recognition of onset consonant changes versus vowel changes in a single familiar word form in British English-learning 12-month-olds

3.3.1 Introduction

The findings from Experiment 2 demonstrated that consonants and vowels appeared to matter equally to British English-learning 11-month-olds when they were presented with lists of mispronounced familiar word forms. However, the absence of a preference for vowel modified words over consonant modified words, as found in French-learning 11-month-olds (Poltrock & Nazzi, 2015), is not proof that infants treat the two conditions similarly. Indeed, the evidence from the Bayesian analysis in Experiments 2b and 2c only provided inconclusive evidence of this null hypothesis. To come to that conclusion, it would need to be demonstrated that, when presented with a correct pronunciation versus a consonant modification on one hand, and with a correct pronunciation versus a vowel modification on the other hand, infants show a preference for the correct pronunciation in both cases equally. Alternatively, if infants show a preference for the correct pronunciation in the consonant modified version only, this would be taken as evidence that they have learned a consonant bias for lexical processing.

Experiment 3 will explore this in a final head-turn preference experiment on a new group of infants. In a between-participants design, infants will be tested on their preference for a single familiar word (*mummy*; /mʌmi/), versus a consonant modified version (*nummy*; /nʌmi/) in one condition, and a vowel modified version (*memmy*; /mɛmi/) in another. Note that this word and the corresponding changes were used in the previous experiments. This task simplification was introduced to ensure the best conditions to observe a consonant bias, if any, at the end of the first year of life. Firstly, this would simplify infants' working memory load by reducing the number of items presented to them. Secondly, a slight increase was made to the ISI from 600ms to 1 second to provide infants with more time to process information. Finally, the age tested was increased from 11 months to 12 months, so that there was potentially a slightly larger receptive lexicon to perform correlation analyses on and allowing for further consolidation of infants' knowledge of their native consonants (Polka & Werker, 1994).

3.3.2 Method

Participants

A new sample of 48 healthy British English-learning 12-month-old infants were tested (mean age = 11;30, range = 11;14 to 13;10, 26 females, 22 males). The data of an additional 21 infants were excluded from the analysis due to crying or being distracted (n = 3), having two consecutive trials with looking times below 2 seconds, or having three or more of such trials in total (n = 10), parental interference (n = 3), technical issues (n = 2) or being an outlier (i.e., the mean orientation times were 2 SDs below or above the group mean; n = 3). The 100-word Oxford Short Form CDI (Floccia et al., 2018; Hamilton et al., 2000) was completed by 47 of the 48 parents as a measure of their infant's vocabulary. All infants were born and raised in the South West of England.

Materials

The stimuli presented to infants was identical to those used in the phonetic conditions in Experiment 1 (see Table 3a for the acoustic features of these tokens).

Procedure

The experiment used the same apparatus and procedure as Experiments 1 and 2 with some minor changes. In a between-participant design, the test phase comprised eight test trials, four of each condition (correct pronunciation lists and consonant or vowel mispronunciation lists). Trials were organised into two blocks that each contained two lists of a correct pronunciation of the word *mummy* and two lists of either the consonant (*nummy*) or vowel (*memmy*) alteration of the same word. Infants were randomly assigned to one of the two conditions, consonant change (n = 24) or vowel change (n = 24). The order the lists were presented in each block was randomised.

3.3.3 Results and discussion

All trials with looking times of under 2 seconds in duration were removed from the analysis. This led to the exclusion of 24 of the 392 test trials (6.12%). All infants included in the analysis provided a minimum of 6 out of 8 useable trials. Mean looking times to the correct pronunciation and altered pronunciation were calculated for each infant. Group averages are presented in Figure 8. Looking times were normally distributed, as revealed by a non-significant Shapiro-Wilk test (W = .96, p = .14 in Correct Pronunciation; W = .99, p = .80 in Altered Pronunciation).

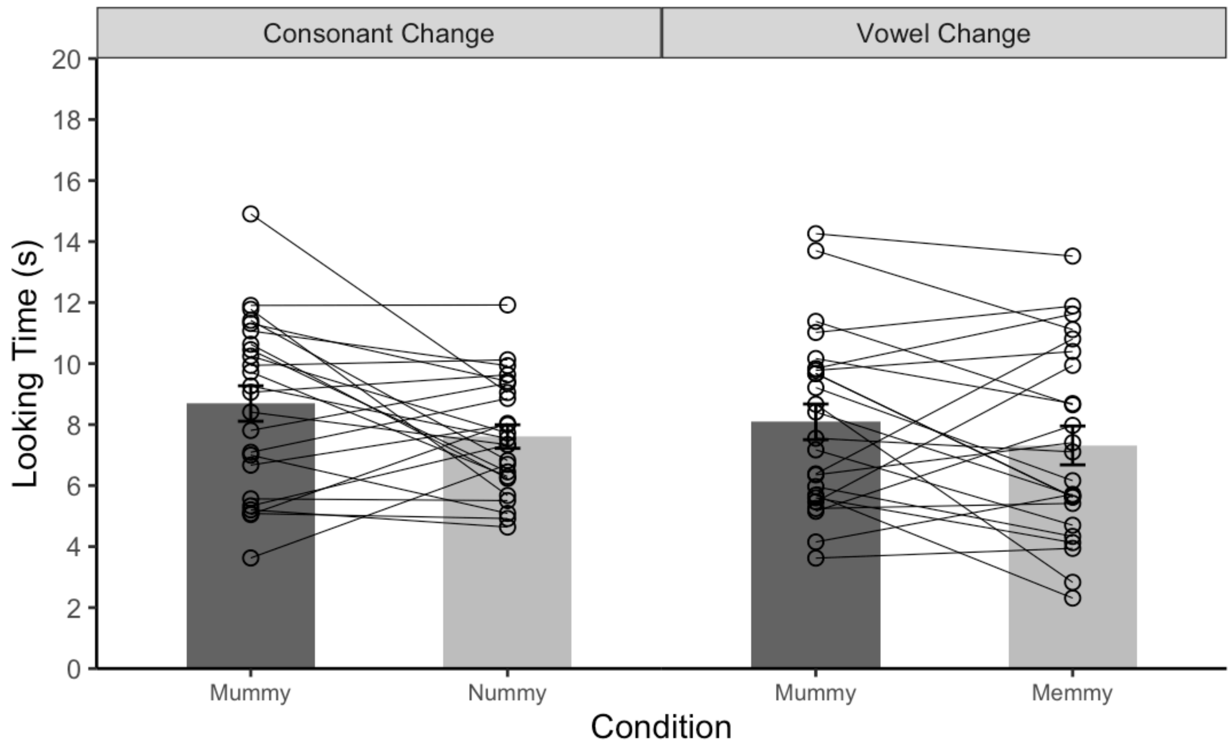


Figure 8: *Mean looking times* (*s*) *in Experiment 3 toward the correct pronunciations of the word mummy versus the altered pronunciations, in the consonant change condition* (*left*) *and vowel change condition* (*right*). *Connected dots represent individual infants' looking times in the two experimental conditions. Error bars represent ±1 standard error.*

A three-way mixed ANOVA was conducted on infants' orientation times, with a within-participant factors of Pronunciation (correct versus altered) and Block (first half versus second half of the experiment) and a between-participant factor of Condition (consonant versus vowel). There was a significant effect of Pronunciation, $F(1, 46)$ =

6.90, $p = .012$, $\eta_G^2 = .02$, but no significant effect of Condition, $F(1, 46) = .35$, $p = .56$, $\eta_G^2 = .003$, or Pronunciation x Condition interaction, $F(1, 46) = .40$, $p = .53$, $\eta_G^2 = .001$.

There was a significant effect of Block, $F(1, 46) = 39.48$, $p < .001$, $\eta_G^2 = .16$, with infants displaying longer orientation times in the first half of the study. However, neither the Block x Pronunciation, $F(1, 46) = .10$, $p = .75$, $\eta_G^2 < .001$, Block x Condition, $F(1, 46) = .34$, $p = .56$, $\eta_G^2 = .002$, nor Block x Pronunciation x Condition, $F(1, 46) = .94$, $p = .34$, $\eta_G^2 = .004$, interactions were significant.

Bayesian statistics were again calculated to estimate the degree of confidence in this null finding (no effect of condition). A Bayes independent samples *t*-test comparing looking times for mispronunciations in the consonant change and vowel change conditions revealed a Bayes factor of BF = .31 ($t(46) = -.39$, $p = .70$). This suggests evidence for the null hypothesis.

Due to the non-significant effect of condition on the preference for correct pronunciations, the impact of lexical factors was calculated across both the consonant change and vowel change conditions. Infants had an average CDI comprehension of 12.57 out of 100 (*SD* = 10.23) words (all children were reported as knowing the word *mummy*). No correlation was found between infants' listening preferences and the CDI ($r(45) = -.10$, $p = .52$, BF = .22), indicating support for the null hypothesis. In relation to consonant and vowel tiers, infants knew an average 12.21 (*SD* = 9.72) unique consonant tiers and 10.21 (*SD* = 6.99) unique vowel tiers. The proportion of unique consonant tiers out of known words (*M* = .99, *SD* = .03) was found to be significantly different to the proportion of vowel tiers (*M* = .90, *SD* = .12), $t(46) = 4.47$, $p < .001$. However, neither proportion of consonant ($r(45) = -.06$, $p = .70$, BF = .20) nor vowel tiers

($r(45) = .07$, $p = .66$, BF = .20) was correlated with infants' listening preferences, again supporting the null hypothesis. No significant differences were found when examining the correlations between infants' listening preferences and CDI, and proportion to either consonant or vowel tiers in the consonant change and vowel change conditions separately ($p > .05$ in all cases, BF = .29 to 1.29).

Results of Experiment 3 show an overall mispronunciation effect, such that infants consistently preferred listening to the correct version of *mummy* over a consonant change or a vowel change. However, as in Experiment 2, there was no evidence of a consonant bias in British English-learning 11-month-olds since preference for the correct version of *mummy* was similar across both consonant and vowel change conditions. These findings again contrast with the preference for vowel mispronunciations over consonant mispronunciation in French (Poltrock & Nazzi, 2015), but align with the absence of a consonant bias in British English word learning (Floccia et al., 2014) and word recognition data (Mani & Plunkett, 2007, 2010).

3.4 Interim discussion

Consonants have been found to be more important than vowels in lexical processing tasks in adults in most languages (for the original proposal, Nespor et al., 2003; for a review, Nazzi & Cutler, 2019). However, rather than being an innate bias, research into the development of this consonant bias suggests that cross-linguistic differences, based on phonological and/or lexical properties of an infants' native language, modulate its acquisition (Nazzi et al., 2016). Chapter Two aimed to explore this further, focussing on the phonetic specificity of British English-learning infants' early word form recognition during their first 12 months. Additionally, it examined whether manipulating the intensity of consonants and vowels impacted word recognition (Experiment 1), and if lexical factors were correlated with the presence of a consonant bias (Experiments 2 and 3).

In the first of a series of head-turn preference studies, Experiment 1 found that British English-learning 5-month-old infants' recognition of the word *mummy* was equally impacted by consonant and vowel changes. This replicated previous research by Delle Luche et al. (2017), who demonstrated that British English-learning infants did not differentiate between consonant and vowel mispronunciations of their own name. This finding, combined with the present results, provides additional evidence to the hypothesis that the consonant bias is learned. Furthermore, it adds additional weight to the argument that cross-linguistic variations exist in the acquisition of the consonant bias. The lack of any lexical processing bias found here is in contrast with French-learning infants, who display a vowel bias in own name recognition at this age (Bouchon et al., 2015). Experiment 1 also investigated whether changes in intensity to either the onset consonant or vowel of the word *mummy* were preferred to unaltered tokens. This was examined based on post-hoc findings in Delle Luche et al. (2017), which suggested that energy cues may impact word form recognition. However, when specifically controlling for this acoustic change, infants did not show any preference. This suggests that 5-month-old infants do appear to use energy to recognise familiar word forms.

It is also important to note that the position of the vowel change in Experiment 1, occurring on the second sound of the word, differed from past research that presented onset vowel changes to infants (Bouchon et al., 2015; Delle Luche et al., 2017; Von Holzen & Nazzi, 2020). However, the vowel change presented to infants still occurred within the initial syllable of the word and thus was likely to still be highly salient to the infants (Jusczyk et al., 1993). The result also replicated the lack of a vowel bias showed in infants presented with an onset vowel change (Delle Luche et al., 2017). Furthermore, past research has shown that French-learning 6-month-olds are also sensitive to vowel changes made to the second sound of a monosyllabic word (Nishibayashi & Nazzi, 2016). Taken together, this suggests that British English-learning infants do not show a vowel bias, irrespective of whether the vowel change is made to the onset of a familiar word or to its second sound.

The results of Experiments 2 and 3 further substantiate the idea that the privileged role of consonants in lexical processing emerges during development. Experiment 2a first showed that 11-month-olds preferred listening to lists of familiar word forms compared to pseudowords. This demonstrates that, by the time infants reach their first birthday, they are beginning to recognise frequently heard word forms in their environment (Hallé & de Boysson-Bardies, 1994; Swingley, 2005; Vihman et al., 2004). Experiments 2b and 2c then found that, at 11 months, infants treated consonant and vowel alterations of the familiar words equally.

Experiment 3 simplified the task for infants, comparing their preference for just one familiar word form and either a consonant or vowel mispronunciation. It also increased the age tested to 12 months. However, infants still did not show a preference for either phonetic alteration. The results reported from Experiments 2 and 3 again diverge from findings in French-learning infants at this age, who show a consonant bias for mispronunciations of lists of familiar word forms (Poltrock & Nazzi, 2015) and their own name (Von Holzen & Nazzi, 2020). Instead, they add support to research showing that British English-learning infants show no preference for either consonants or vowels at 12, 18, and 24 months in a word recognition task (Mani & Plunkett, 2007, 2010) and at 24 months in a word-learning task (Floccia et al., 2014).

Experiments 2 and 3 additionally measured the relationship between vocabulary size, along with the number of consonant/vowel tiers they understood, and infants' preferences for consonant or vowel mispronunciations. The results revealed that no correlation existed between the lexical measures taken and infants' preference for either type of phoneme change. This replicates findings in French-learning infants, where lexical measures are not associated with infants showing a consonant bias in familiar word (Poltrock & Nazzi, 2015) and own name (Von Holzen & Nazzi, 2020) recognition. Thus, it appears that lexical factors, at least as measured by the number of words and consonant vowel tiers infants know, may not account for the presence of a consonant bias at 11 and 12 months.

4 Chapter Three

4.1 Introduction

The research reviewed in Chapter One, and the experiments reported in Chapter Two, primarily focused on infants' auditory speech perception. However, in everyday communication, language does not typically occur in isolation. Instead, speech is a multimodal process (Rosenblum, 2008), perceived through highly correlated and temporally coordinated auditory (i.e., through hearing) and visual (i.e., through articulatory face movements) signals (Chandrasekaran et al., 2009). Observing audiovisual speech provides additional information concerning both the phonetic and temporal properties found in the acoustic signal (Peelle & Sommers, 2015). As will be discussed in detail below, both adult (e.g., Fort et al., 2010) and infant (e.g., Teinonen et al., 2008) speech perception benefits from the concurrent presentation of such audio and visual cues. Given that the British English-learning infants tested in Chapter Two struggled to identify phonetic changes to familiar word forms, providing them with the supplementary phonetic cues afforded by audiovisual speech could potentially assist them with recognising consonant and vowel mispronunciations. This will be explored in the following two chapters. Chapter Three will first describe the impact of audiovisual cues on speech perception in adults, whether infants can match audio and visual signals, and infants' face scanning behaviour when viewing speech, and detail how access to audiovisual stimuli can enhance speech perception in infants. Chapter Four will then present research examining the impact of audiovisual speech cues on British English-learning infants' capacity to recognise phonetic mispronunciations of familiar word forms.

4.2 Audiovisual influences on speech perception in adults

When adults view a face speaking in their native language they will typically focus on the talker's eyes (Lewkowicz & Hansen-Tift, 2012; Morin-Lessard et al., 2019; Vatikiotis-Bateson et al., 1998). Attending to an individual's eyes is extremely beneficial to adults, providing a range of communicative signals such as where to direct their attention (Ricciardelli et al., 2002), and inferring another's mental state (Baron-Cohen et al., 2001). However, adults will often utilise visual information from the mouth if speech processing becomes challenging. For instance, it has long been established that audiovisual cues can enhance adults' abilities to successfully perceive and discriminate speech when noise negatively impacts the acoustic information in an auditory signal (Sumby & Pollack, 1954). In such noisy scenarios, access to the articulatory movements on a speaker's face improves the comprehension of phonemes (Fort et al., 2010), words (Ross et al., 2007), and sentences (Grant & Seitz, 2000). To illustrate, Fort et al. (2010) asked adults to respond when they heard a phoneme contained in a series of words and pseudowords that were presented in white noise. Participants were both faster and more accurate at correctly identifying the phoneme when the target word was presented audiovisually rather than auditorily. Furthermore, this effect was stronger for words than pseudowords, indicating that viewing the phonetic information provided by the mouth led to lexical activation. Fixations on the mouth are found to significantly increase when the listening environment is noisy, indicating that adults will preferentially seek out visual information when the auditory signal is harder to perceive (Buchan et al., 2008; Vatikiotis-Bateson et al., 1998; Yi et al., 2013). Accessing such audiovisual information can significantly reduce the cognitive effort required to perceive the speech signal in noise (Fraser et al., 2010).

Adults also benefit from access to a speaker's mouth when discriminating between difficult non-native phonemic contrasts. For instance, training with audiovisual cues, as opposed to auditory-only, during a learning phase improved Japanese adults' auditory perception of non-native English consonants (Hazan et al., 2005) and English adults' auditory perception of Japanese vowel length contrasts (Hirata & Kelly, 2010) in a later test phase. Adults perform better when segmenting words from an artificial language when they are provided with audiovisual cues, rather than audio or visual cues alone (Mitchel & Weiss, 2014). Relatedly, monolinguals increase their fixations on a non-native talker's mouth but not a native talker's mouth during complex speech processing tasks (Barenholtz et al., 2016). Even highly proficient second language speakers attend more to the mouth when hearing their non-native second language (Birulés et al., 2020). Therefore, audiovisual cues appear to be advantageous to adults when they are processing unfamiliar speech.

Behavioural and neurological research has also shown that audiovisual speech can trigger lexical access and word recognition even in normal listening conditions. For example, Kim et al. (2004) reported that adults were faster at a lexical decision task when they were primed with silent visual articulations of target words. This effect was only observed for real words and not pseudowords, which again suggests that visual speech impacts lexical retrieval. In an ERP experiment, Kaganovich et al. (2016) found that the incongruent presentation of auditory-only words followed by visual-only words resulted in a higher MMN response than a congruent presentation. Matching audio and visual pairings also elicited a higher late positive complex, an ERP component that is sensitive to word repetition (Rugg & Curran, 2007). Even viewing the silent visual articulation of the initial syllable of a word can facilitate lexical access. Fort et al. (2013) discovered that French adults were faster at recognising audio-only words (e.g., /bonɛ/, meaning *cap*) when they were preceded by a visual-only presentation of their onset syllable (e.g., /bo/) compared to an unrelated syllable (e.g., /vu/). Furthermore, viewing the mouth improves the comprehension of semantically complex passages and accented speech when presented audiovisually rather than just auditorily (Arnold & Hill, 2001; Reisberg et al., 1987). Overall, the cortical processing of auditory speech is faster when presented audiovisually than auditorily (Skipper et al., 2007; van Wassenhove et al., 2005). Together, these findings demonstrate that the multimodal nature of speech can be extremely beneficial to adults during lexical processing tasks.

A further demonstration of how viewing a speaker's mouth can influence speech perception in adults is the McGurk effect (McGurk & MacDonald, 1976). This illusion presents an auditory syllable alongside the incongruent visual articulation of a different syllable. Specific syllable pairings can lead to the integration of audiovisual information into a unitary percept. To illustrate, an audio /ba/ and visual /ga/ often results in a fused percept of either /da/ or /ðɑ/. In contrast, the reverse pairing can lead to non-fusible percept /bga/, which is a phonetically illegal phoneme in the English language (McGurk & MacDonald, 1976). Other mismatching audio and visual syllables can lead to the visual signal being perceived over the auditory signal. For instance, an audio /ba/ paired with a visual /va/ is typically heard as /va/, whereas an audio /da/ paired with a visual /va/ is heard as /da/ (Rosenblum & Saldaña, 1992, 1996). Research examining the neural basis of the McGurk effect has revealed that activity in the auditory cortex is influenced by visual speech. Participants who have been familiarised to an audio /ba/ and visual /ba/ show a

MMN response when the visual signal changes to /va/, even though the auditory signal has remained the same (Saint-Amour et al., 2007). Whilst there is evidence of a correlation between mouth fixations and the magnitude of the McGurk effect (Gurler et al., 2015), the perception of the illusion remains relatively stable even when instructing adults to direct their attention towards other areas of the face (Paré et al., 2003).

4.3 Mechanisms underlying audiovisual speech benefits in adults

Several factors have been proposed to explain why speech perception is positively influenced when concurrent audio and visual signals are presented to a listener (for a review, see Lalonde & Werner, 2021). Firstly, the movements of the mouth, lips, and teeth can supply useful phonetic details that are not supplied by the auditory signal alone. These additional cues may assist adults with discriminating between different phonemes (e.g., Lalonde & Holt, 2015; Owens & Blazek, 1985). For consonants, whilst both audio and visual signals can communicate information relating to their manner of articulation, and the audio signal conveys cues such as their voicing and nasality, the visual signal alone details their place of articulation (Ladefoged & Disner, 2012). Indeed, cues relating to the manner of articulation (e.g., (e.g., /b/ versus /v/), as well as cues such as voicing (e.g., /d/ versus /t/), are easier to distinguish auditorily than visually (Summerfield, 1987), whereas cues concerning the place of articulation (e.g., /b/ versus /d/ versus /g/) are easier to distinguish visually than auditorily (Miller & Nicely, 1955). Vowels are less visually recognisable than consonants, given that they do not differ in their manner or place of articulation, but rather in the position of the tongue (Stevens & House, 1955). Nevertheless, the mouth becomes more open as the height of the tongue lowers, such that /a/ has a visibly more open jaw than /i/. Visible lip rounding in the production of back vowels will also reduce as the tongue height lowers, with /u/ being visibly more rounded than /ɔ/ (Ladefoged & Disner, 2012). Diphthongs are also typically visually more distinctive compared to monophthongs (Wozniak-Kaelin & Jackson, 1979). However, the visual signal overall has been found to typically provide more information for discriminating between vowel roundness, followed by height, and then backness (Benguerel & Pichora-Fuller, 1982; Robert-Ribes et al., 1998).

Secondly, audiovisually presented speech may influence lexical activation. As previously mentioned, visual articulations appear to specifically impact word recognition both in noise (Fort et al., 2010) and in normal listening conditions (Kim et al., 2004). Similarly, audiovisual speech may impact the activation of lexical neighbourhoods (Tye-Murray et al., 2007), as depicted in Figure 9. When an auditorily presented word is heard (e.g., *fork*), it is thought to activate a neighbourhood of phonologically similar sounding words that differ by either the substitution (e.g., *fort)*, deletion (e.g., *for*), or addition (e.g., *forked*) of a single phoneme (Luce & Pisoni, 1998). A spoken word with a sparse auditory neighbourhood is usually perceived more accurately and quickly than one with a dense auditory neighbourhood (Chan & Vitevitch, 2015; Luce & Pisoni, 1998; Vitevitch & Luce, 1998). An equivalent visually defined lexical neighbourhood has also been proposed (Mattys et al., 2002), in which a visually presented word activates other words which share comparable facial and articulator movements (e.g., *force, ford, fort*). As with auditory neighbourhoods, words with sparse visual neighbourhoods are more likely to be recognised (Auer, 2002; Mattys et al., 2002; Tye-Murray et al., 2007). Importantly, there is evidence that the overlap created by the simultaneous activation of auditory and visual neighbourhoods can decrease the number of competing word candidates and improve word recognition (Feld & Sommers, 2011; Tye-Murray et al., 2007). For instance, whilst *fork* (see Figure 9a) has a dense overlap when it is audiovisually presented (e.g., *force*, *fort*, *fort*, etc.), the word *fish* (see Figure 9b) has no words which are both auditorily and visually comparable to it. Thus, an audiovisually presented *fish* is more likely to be correctly identified than either an audio-only or visual-only presentation. Therefore, audiovisual speech may be beneficial in helping to constrain lexical selection, which consequently may lead to correctly identifying the speech signal (Tye-Murray et al., 2007).

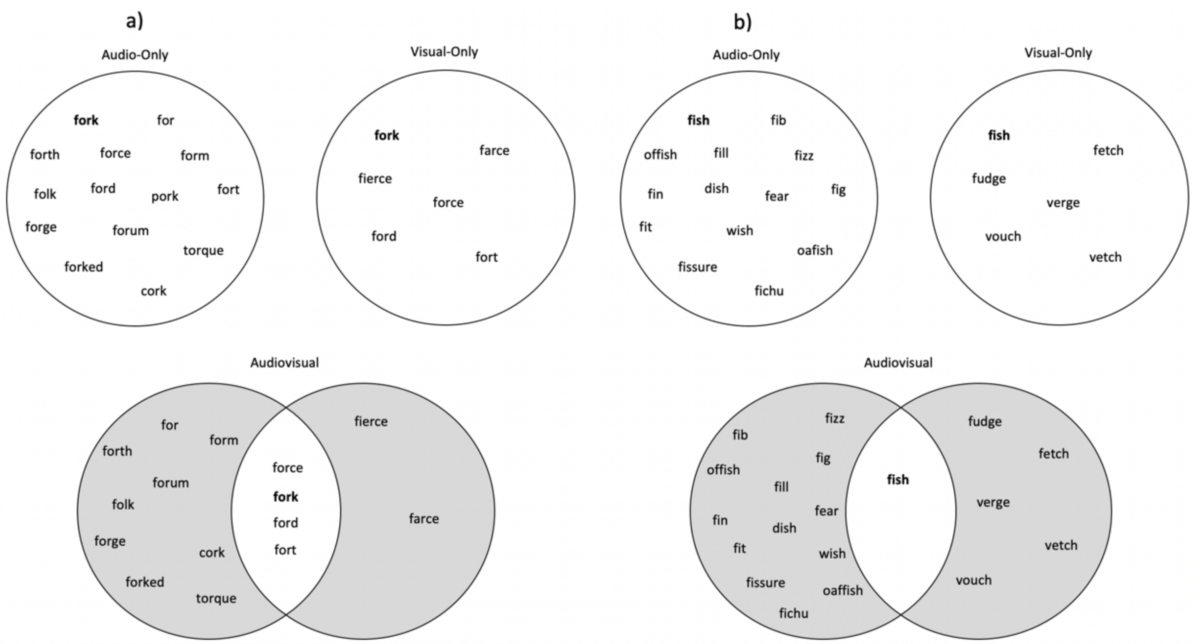


Figure 9: *Audio-only, visual-only, and audiovisual lexical neighbourhoods for the words a) fork and b) fish (adapted from Tye-Murray et al., 2007).*

Thirdly, the temporal relationship between audio and visual speech signals may also strengthen speech perception. The onset cues of the audio and visual signal are highly correlated, with mouth movements occurring between 60 and 300ms before the voice is audible (Chandrasekaran et al., 2009; Schwartz & Savariaux, 2014). Similarly, the amplitude envelope of the auditory signal is congruent with visual articulations, with speech becoming louder as the mouth opens and quieter as the mouth closes (Chandrasekaran et al., 2009). Adults are sensitive to the synchrony of audio and visual speech, detecting asynchrony when the auditory signal precedes the visual signal by between 60 to 200ms and when the visual signal precedes the auditory signal by between 180 to 240ms (e.g., Dixon & Spitz, 1980; Grant et al., 2004; van Wassenhove et al., 2007). Such congruent audio and visual cues improve temporal expectancy, assisting listeners with when to direct their attention towards the auditory input. For example, Grant and Seitz (2000) found that adults' ability to detect the presence of sentences masked by white noise was significantly improved when presented alongside matching visual information compared to mismatching visual information. Overall, the auditory threshold to detect speech was decreased by 1 to 2dB when the audio signal was presented alongside matching articulations. Furthermore, providing adults with animated non-face stimuli (e.g., horizontal ovals) that parallel mouth movements (Tye-Murray et al., 2011), or the amplitude envelope of the speech signal (Yuan et al., 2020), can lead to significant improvements in speech perception, particularly in noisy environments, compared to the auditory signal alone.

To summarise, adults' speech perception is impacted by the concurrent presentation of audio and visual speech signals in a range of tasks. The additional phonetic cues (Lalonde & Holt, 2015), lexical activation (Tye-Murray et al., 2007), and temporal information (Grant & Seitz, 2000) provided by viewing a speaker's mouth have been proposed as explanations for such an audiovisual benefit. The following section will detail whether infants are susceptible to such phonetic and temporal cues, before examining whether their speech perception during the first year of life is improved when the audio and visual signals are simultaneously presented to them.

4.4 Audiovisual speech perception in infants

Infants are frequently exposed to faces during their first year of life (Fausey et al., 2016; Sugden et al., 2014). From birth, they show an attentional bias towards looking at faces in their environment (Cassia et al., 2004; Johnson et al., 1991), a preference which becomes more robust and advanced during their first 6 months (Leppänen, 2016). Perhaps unsurprisingly, given infants' early interest for auditory speech (Vouloumanos & Werker, 2007), a talking face becomes increasingly salient to infants over the course of their first year, such that by 6 to 8 months they will fixate longer on a speaking face than either a silent face or other audiovisual objects (Bahrick et al., 2016).

Beyond preferentially attending to the audiovisual presentation of a face, infants show a remarkable ability to detect the correspondence between speech sounds and their associated mouth articulations from very early in ontogeny. Kuhl and Meltzoff (1982) presented 4.5-month-olds with an intersensory matching procedure, in which they were shown two simultaneous side-by-side videos of the same face, one silently articulating /a/ and the other silently articulating /i/. An audio /a/ or /i/ was presented in synchrony with the articulatory movements of each face but was crucially congruent with only one face. Infants were found to look significantly longer at the face that matched the auditory stimuli. This indicates that, even at this young age, infants perceived the cross-modal congruence between audio and visual speech cues. This result has subsequently been replicated in

2-month-old infants (Patterson & Werker, 2003), in newborns using a simplified procedure (Aldridge et al., 1999), and with consonant-vowel disyllables (e.g., /zuzu/) in 5- to 6-month-olds (MacKain et al., 1983). However, some native phonetic contrasts may be harder to match than others based upon their visual distinctiveness. For instance, Altvater-Mackensen et al. (2016) found that German-learning 6-month-olds could successfully match the more visually distinct /a/ and /o/ but not the less visually distinct /a/ and /e/. For continuous speech, infants aged 2.5 to 5 months prefer a face reciting a nursery rhyme in their native language when it is audiovisually synchronised rather than to desynchronised by 400ms (Dodd, 1979). Similarly, 3- and 7-month-olds, although not 5-month-olds, look significantly longer towards the synchronous video that matches native fluent speech in a simultaneous intersensory matching procedure (Pickens et al., 1994). Infants can also pair a previously familiarised auditory phoneme or fluent speech to their correct silent articulation in a sequential intersensory matching task (e.g., Kubicek et al., 2014; Pons et al., 2009). This suggests that temporal synchrony is not critical for matching audio and visual speech signals.

Additionally, infants can pair non-native sound contrasts or speech to seen articulations during the early stages of development. As with perceptual attunement of auditory phonemes, such audiovisual matching narrows towards an infant's native language during their first 12 months (e.g., Danielson et al., 2017; Dorn et al., 2021; Kubicek et al., 2014; Lewkowicz & Pons, 2013; Pons et al., 2009). For phonemes, Pons et al. (2009) tested English- and Spanish-learning 6- and 11-month-olds' ability to discriminate a /b/-/v/ contrast, which is found in English but not Spanish. In a sequential intersensory matching task, infants were first auditorily familiarised with one of the test phonemes, before viewing side-by-side videos of a female speaker silently uttering each phoneme. At 6 months, both English- and Spanish-learning infants matched the previously heard phoneme to the correct face. In contrast, at 11 months, only English-learning infants were successful at the task, with Spanish-learning infants no longer able to match audible to visible non-native phonemes. For native and non-native speech, Kubicek et al. (2014) found that 4.5-month-olds could sequentially match native (German) and non-native (French) passages of speech to the correct silent visual articulations in a sequential intersensory matching task. However, the ability to match native language over non-native language may appear later, at 10 to 12 months, when the audio and visual cues are presented simultaneously rather than sequentially (Lewkowicz & Pons, 2013). Nevertheless, these findings suggest that infants' audiovisual perception of phonemes and fluent speech has narrowed to their native language by their first birthday.

Infants also appear to successfully integrate audiovisual cues. For instance, although the McGurk effect has been found to be weaker during the first year of life (Desjardins & Werker, 2004), with the illusion's impact on auditory perception gradually increasing during childhood (McGurk & MacDonald, 1976; Sekiyama & Burnham, 2008), certain audio and visual syllable pairings can impact what an infant perceives. Burnham and Dodd (2004) habituated 4.5-month-olds to an incongruent audio /ba/ and visual /ga/, which adults often perceive as the fused percept of /da/, or /ðɑ/ (McGurk & MacDonald, 1976). They then tested whether infants dishabituated (i.e., looked longer) when hearing an auditory-only /ba/, /da/, or /ðɑ/. Dishabituation in this context would indicate that infants had detected a change between a familiarised and a newly presented stimulus. Infants dishabituated to /ba/ but not to either /da/ or /ðɑ/. The authors proposed that infants had not perceived the audio /ba/ during habituation, leading them to dishabituate to the audio /ba/ at test. Instead, infants, just like adults, had integrated the audiovisual speech cues into a fused /da/ or /ðɑ/ percept, which explained their failure to dishabituate when hearing these syllables.

Relatedly, Rosenblum et al. (1997) habituated 5-month-olds with a matching audio /va/ and visual /va/, before changing the audio stream to either an incongruent /ba/ or /da/. Recall that adults perceive these mismatched pairings as /va/ (i.e., visual over auditory signal) and /da/ (auditory over visual signal) respectively (Rosenblum & Saldaña, 1992, 1996). When the audio switched to /ba/, infants did not dishabituate, suggesting that they too continued to perceive the visual instead of the audio signal. However, infants did dishabituate when the audio changed to /da/. This indicates that infants, like adults, perceived the audio rather than the visual signal for this incongruent pairing. Thus, certain visual cues can alter how an infant perceives an auditory signal.

Furthermore, brain imaging studies have revealed that infants integrate audio and visual information. Bristow et al. (2008) measured 10-week-old infants' ERPs for a mismatch response when given a cross-modal-matching task. Trials were either the familiarisation to an audio vowel (e.g., /i/) before the viewing of a silent visual articulation of a congruent or incongruent vowel (e.g., /a/), or a reversed visual then audio presentation. Incongruent vowels elicited a MMN response in infants, a finding that was irrespective of whether familiarisation had been auditory or visual. This result aligns with behavioural evidence (e.g., Kuhl & Meltzoff, 1982) and demonstrates that the neural systems involved in audiovisual speech processing are present in infants. In another ERP study, Kushnerenko et al. (2008) measured 5-month-olds' mismatch response to the

McGurk effect. Infants did not show a MMN response to an audio /ba/ and visual /ga/, indicating that they integrated the audiovisual cues into an adult-like fused /da/ or /ðɑ/. However, when presented with an audio /ga/ and visual /ba/, which leads to a non-fusible /bga/ in adults (McGurk & MacDonald, 1976), infants displayed a MMN response. This suggests that infants also perceived the phonetically illegal /bga/ and recognised the incongruence between the audio and visual signals.

Finally, it is important to acknowledge that it remains unclear whether the phonetic and/or temporal cues provided by viewing a speaking face account for audiovisual speech matching during an infant's first 12 months (for a review, see Lalonde & Werner, 2021). Infants are clearly sensitive to audiovisually presented phonetic cues, as demonstrated by their ability to match concurrently presented auditory and visual signals (Aldridge et al., 1999; Kuhl & Meltzoff, 1982; Patterson & Werker, 2003) and detect audio and visual phonetic mismatches (e.g., Kushnerenko et al., 2008; Tomalski et al., 2013). Infants can also pair audio and visual phonemes sequentially, demonstrating that temporal synchrony is not critical for such audiovisual matching (Pons et al., 2009). Similarly, as detailed below, audiovisual stimuli can impact the discrimination of both native and non-native phonemes (Teinonen et al., 2008; Ter Schure et al., 2016) and mispronunciation detection (Weatherhead & White, 2017).

However, infants' early speech perception may rely more on temporal than phonetic information. Neurologically, the cortical mechanisms involved in detecting temporal synchrony have developed by 6 months, whereas phonetic/lexical representations show limited maturation during an infant's first 12 months (Eggermont & Moore, 2012). Behaviourally, infants, but not adults, can match audio and visual signals based upon temporal information alone. For example, Baart et al. (2014) tested adults' and 5- to 15-month-olds' abilities to pair trisyllabic pseudowords (e.g., *kalisu*), presented in unprocessed or sine-wave speech (SWS), to their correct visual signal. SWS maintains the temporal information of the speech signal, but the phonetic information is significantly compromised. Adults' matching performance was significantly poorer for SWS compared to unprocessed speech. In contrast, infants were found to successfully pair both unprocessed speech and SWS to its corresponding visual signal in an IPL task. Similarly, infants' speech perception can be improved when they are provided with only temporal information contained in the speech signal (Hollich et al., 2005; Lalonde & Werner, 2019). This demonstrates that temporal, rather than phonetic, information may be more important for infants when detecting the correspondence between speech sounds and their associated mouth articulations.

4.5 Benefits of audiovisual speech in the first year of life

Although under researched in comparison to the adult literature, there is evidence that infants' speech perception also benefits from the concurrent presentation of audio and visual signals (Teinonen et al., 2008; Ter Schure et al., 2016; Weatherhead & White, 2017). As will be discussed in detail below, audiovisual speech has been found to impact distributional learning (Teinonen et al., 2008), the acquisition of non-native phonetic contrasts following perceptual attunement (Ter Schure et al., 2016), syllable detection in noise (Lalonde & Werner, 2019), word segmentation (Hollich et al., 2005), and mispronunciation detection (Weatherhead & White, 2017).

4.5.1 Distributional learning

The audiovisual enhancement of speech perception in the first year of life has been demonstrated in distributional learning, whereby infants have been found to use frequency distributions in a continuous auditory signal to discriminate between different speech sounds (Maye et al., 2002). Distributional learning experiments typically familiarise infants with an equally spaced continuum of sounds that form a phonological contrast (e.g., /ba/-/da/), before testing their ability to discriminate between the two endpoints. However, whilst tokens at the two endpoints of the continuum are equally presented during familiarisation, the distribution of the sounds along the continuum differs, being either bimodally distributed (i.e., sounds towards the two endpoints being more frequent), or unimodally distributed (i.e., sounds from the middle of the continuum being more frequent).

Auditory-only studies have revealed that familiarisation to either a bimodal or unimodal distribution differentially impacts infants' subsequent discrimination (Liu & Kager, 2014; Maye et al., 2002, 2008; Wanrooij et al., 2014; Yoshida et al., 2010). For example, Maye et al. (2002) familiarised English-learning 6- and 8-month-old infants with eight equally spaced sounds from a native /da/- /ta/ continuum, which were either bimodally or unimodally distributed. In a test phase, the infants' ability to discriminate tokens from the endpoints of the continuum was measured using the stimulus alternation preference procedure (Best & Jones, 1998). This paradigm presents infants with trials that are either an alternating stimulus (e.g., /da/ /ta/ /da/ /ta/…) or a non-alternating stimulus (e.g., /da/ /da/ /da/ /da/…). Given that discrimination between the two trial types requires infants to perceive variations, differential attention to either alternating or non-alternating trials indicates that infants have detected a change. Using this method, Maye et al. (2002) found that, for both age groups, only infants who had been familiarised with the bimodal frequency were successful in discriminating between the endpoints of the continuum. Although less effective following the acquisition of native phonetic categories (Yoshida et al., 2010), exposure to an auditory bimodal, but not unimodal, frequency distribution can also help infants to discriminate between native and non-native contrasts (Liu & Kager, 2014; Maye et al., 2008; Wanrooij et al., 2014).

However, there is evidence that audiovisually presented unimodal distributions can help infants to differentiate between phonemes. Teinonen et al. (2008) presented English-learning 6-month-olds with a continuum of speech sounds from a native phonological contrast /ba/-/da/. The distribution of the sounds presented was only unimodal, with most tokens occurring around the phonemic boundary between /ba/ and /da/. Alongside the auditory stimuli, infants were presented with synchronous visual speech cues. In a two-category group, sounds nearer to /ba/ on the continuum were paired with a face articulating /ba/, whereas sounds nearer to /da/ on the continuum were paired with a face articulating /da/. In a one-category group, infants viewed a single visual articulation of either /ba/ or /da/, irrespective of where the accompanying sound occurred on the continuum. Following familiarisation, infants' auditory discrimination was tested using the stimulus alternation preference procedure. Infants in the two-category group were found to differentiate between auditory-only /ba/ and /da/ tokens, whereas infants in the one-category group did not. Thus, although infants fail to discriminate auditory-only unimodal phoneme contrasts (Maye et al., 2002), providing matching articulatory cues appeared to assist their ability to recognise phonemes.

4.5.2 Phonetic discrimination

Research has also examined whether audiovisual cues may help infants to learn non-native phonetic contrasts beyond perceptual attunement. Ter Schure et al. (2016) examined how bimodal and unimodal distributions shaped Dutch-learning 8-month-olds' recognition of a non-native vowel /æ/-/ɛ/ contrast when they were combined with different audio and visual information. This phonetic contrast is found in British English but not in Dutch, with Dutch adults instead mapping /æ/-/ɛ/ onto the single Dutch phoneme /ɛ/ (Weber & Cutler, 2004). By 8 months, Dutch infants' vowel perception is likely to have narrowed to their native phonetic categories (Kuhl et al., 1992; Polka & Werker, 1994), so they would potentially struggle to differentiate between this non-native contrast. Across three different conditions, infants were familiarised with a face articulating either a bimodal or unimodal distributed /æ/-/ɛ/ contrast, presented in a /f_p/ consonant context (e.g., /fæp/ and /fɛp/). In an audiovisual condition, infants viewed the face articulating the contrast with matching audio. In an audio condition, the speaker's mouth was obscured with their hand, such that no articulatory information was conveyed. Finally, in a video condition, the articulatory cues were presented but in the absence of any auditory information. The results showed that only infants who had been presented with both a bimodal frequency and audiovisual cues were successful in discriminating the non-native contrast at test. This suggests that observing both audio and visual signals helped infants to discriminate between non-native vowel contrasts. However, Danielson et al. (2017) found that familiarising Canadian English-learning 9-, and 11-month-olds with audiovisually presented non-native Hindi /d̪a:/ and /ɖa:/ contrasts did not enhance their ability to discriminate between auditory-only versions of the same tokens. Therefore, it remains unclear how beneficial audiovisual cues are for improving phonetic discrimination of non-native phonemes beyond perceptual attunement.

4.5.3 Speech in noise

There is evidence that the temporal cues provided by a speaker's mouth can improve infants' speech perception in noise. Lalonde and Werner (2019) examined whether 6- to 8.5-month-olds' and adults' detection and discrimination of syllables in noise (e.g., /mu/) was improved when they were accompanied by visual signals. One group viewed a concurrent visual signal (i.e., a video of the speaker), providing both phonetic and temporal information of the syllable. Another group viewed two still images of the speaker, providing only temporal information of the syllable's onset (a picture of the speaker with their mouth open) and offset (a picture of the speaker with their mouth closed). Overall, the detection of syllables in noise by infants and adults was improved when given either type of visual cues compared to audio alone, with no difference in benefit found between the video of the speaker or the onset-offset images. This demonstrated that onset-offset cues and access to the speaker's articulations equally impacted infants' and adults' speech detection. Likewise, both visual cues impacted infants' and adults' syllable discrimination. However, adults' performance was better when provided with the concurrent visual signal in comparison to the visual onset-offset. In contrast, infants showed no difference in benefit between the two visual cues. Thus, only adults appeared to be utilising the phonetic cues provided by viewing the speaker articulate the syllables, with infants instead relying on the temporal onset-offset cues provided in both types of visual signal to discriminate between syllables. Therefore, it appears that infants may particularly benefit from temporal, rather than phonetic, information when discriminating speech in noise.

4.5.4 Speech segmentation

The benefits of concurrently presented audio and visual signals have also been investigated in relation to infants' speech segmentation. Hollich et al. (2005) examined whether 7.5-month-olds could segment words from a short story when it was concurrently played with a distractor passage. Infants were first familiarised with the auditory stimuli, which was heard alongside either a congruent visual signal, an incongruent visual signal, a still image of the speaker, or an oscilloscope pattern displaying only the temporal envelope of the audio signal. Infants' recognition of a target word contained in the short story was then tested in a head-turn preference task. The results showed that infants presented with the congruent audio and visual signals segmented the target word from the speech signal, whereas infants presented with the incongruent visual signal, or the still image, failed to do so. Interestingly, infants who were provided with the oscilloscope pattern also appeared to successfully segment the target word in the test phase. This suggests that the temporal relationship between the audio and visual signals was particularly beneficial to infants' abilities to segment speech in noise. More recently, Tan et al. (2023) found that, whilst 7-month-olds were successful at segmenting words from passages of fluent speech when provided with either auditory-only or audiovisual speech, infants who viewed a speaker's face showed more persistent segmentation abilities across test trials. Together, these findings demonstrate that speech segmentation in infants benefits from audiovisual speech in both noisy and normal listening conditions.

4.5.5 Word recognition

Infants' detection of mispronunciations of familiar word forms may also be influenced by the presence of visual speech cues. In the first of a series of experiments, Weatherhead and White (2017) presented English-learning 12- to 13-month-olds with videos of a speaker producing a series of separate lists of correctly pronounced familiar word forms (e.g., *baby*, *diaper*, *telephone*) and pseudowords (e.g., *boli*, *dimper*, *tolempill*). Each video was presented on the screen for as long as the infant fixated towards it. During the first half of the study's trials, infants were found to look longer at the speaker when they articulated unfamiliar words. In contrast, during the second half of trials, infants looked equally at the speaker in both conditions. Thus, infants differentiated between the two types of lists presented. Such a novelty preference for unfamiliar words is the opposite to the familiarity preference found in auditory word form recognition studies at a similar age (e.g., Hallé & de Boysson-Bardies, 1994; Vihman et al., 2004). The authors speculated that this difference may be explained by infants increasing their attention towards visual cues, especially the mouth, when hearing novel words to assist with their word learning. However, given that this study did not measure infants' face scanning behaviour, it is not possible to confirm this hypothesis.

A second experiment used the same procedure to examine whether infants discriminated between the familiar and unfamiliar word forms when they were mispronounced at word onset. In one condition, the initial phonemes of the familiar word forms were altered by a voiced change, such that the mispronunciations were visually consistent with the correct pronunciations (e.g., *paby*, *taiper*, *delephone*). In another condition, the initial phonemes of the familiar word forms were altered by a place change (e.g., *daby*, *biaper*, *pelephone*), such that the mispronunciations were visually inconsistent with the correct pronunciations. Infants were found to look longer at the speaker when they produced unfamiliar words, but only when the familiar words were mispronounced with a voiced change. This result was again only observed during the first half of trials. Therefore, infants who viewed the visually consistent mispronunciations successfully differentiated between the familiar word forms and pseudowords. In contrast, infants who had viewed the visually inconsistent mispronunciations did not discriminate between the two types of stimuli. Given that infants had previously shown a preference for novel over familiar word forms, this finding suggests that infants treated visually consistent mispronunciations as familiar word forms and visually inconsistent mispronunciations were treated as unfamiliar word forms.

A final experiment then examined infants' preferences for the same voiced mispronunciations of familiar words and unfamiliar words. However, this time infants were only provided with the audio signal, the presentation of which was congruent on their attention to a checkerboard on a screen rather than an articulating face. In the absence of any visual cues, infants listened equally to each audio-only list. This suggests that the preference for unfamiliar word forms compared to familiar word forms found during the audiovisual presentation was influenced by the presence of the speaker's articulatory features. This importantly demonstrates that infants only treated voiced mispronunciations as familiar word forms when they were presented alongside visual cues.

Weatherhead and White (2017) proposed three factors that may explain why observing visual articulations impacted infants' familiar word form recognition. Firstly, they suggested that viewing the visual onset phoneme of a word may activate comparable phonemes. To illustrate, the phonemes /t/ and /d/ are visually consistent, such that an audiovisual presentation of /t/ will still activate the visual signal of /d/ even though the audio signal does not match. Thus, seeing the initial phoneme /t/ followed by *-iaper* may lead to an infant identifying the word *diaper*. Secondly, viewing the onset visual phoneme of a word may potentially activate the lexical neighbourhood of visually consistent articulations. Indeed, lexical access can be improved in adults when they are primed with the initial articulatory gestures of words (Fort et al., 2013). Therefore, viewing the visual onset phoneme /t/ may activate words which share visually consistent onsets (e.g., *daddy*, *diaper*, *tickle*, *tummy*), leading to the identification of the word *diaper* when audiovisually presented with *tiaper*. Thirdly, the visual signal may also be implemented post-lexically. For instance, whilst both *tiaper* and *biaper* may activate the lexical representation of *diaper*, only *tiaper* is visually congruent with *diaper*. Consequently, the consistency between the activated word and visual articulation may influence word recognition.

4.5.6 Summary

To summarise, research suggests that infants' speech perception, like adults', can be positively impacted by the concurrent presentation of audio and visual speech cues. This can range from improving their abilities to learn unimodally distributed syllables (Teinonen et al., 2008), identifying non-native phonemes (Ter Schure et al., 2016), segment speech (Hollich et al., 2005), and detect mispronunciations (Weatherhead & White, 2017). Whilst it is unclear whether the temporal or phonetic information provided by the speech signal assists speech perception, infants appear to benefit from audiovisual speech in comparison to audio speech alone during their first year of life.

4.6 Developmental trajectory of infant attention to faces articulating fluent speech

Evidence that infants match (e.g., Kuhl & Meltzoff, 1982), integrate (e.g., Burnham & Dodd, 2004), and benefit from (e.g., Teinonen et al., 2008) the concurrent presentation of audio and visual speech cues suggests that they are sensitive to audiovisual speech. Research over the past decade has also begun to explore whether infants selectively attend to either the eyes or a mouth when viewing a talking face, and whether increased speech processing demands influence these looking behaviours. In a seminal study, Lewkowicz and Hansen-Tift (2012) examined visual attention to facial regions in American English-learning 4-, 6-, 8-, 10-, and 12-month-olds, as well as American-speaking adults, whilst they viewed a talking face. Participants were shown videos of a female reciting a monologue in either English or Spanish, whilst an eye tracker measured fixations towards the speaker's eyes, mouth, and face. The proportion of total looking time to the eyes and mouth was calculated by dividing the number of fixations on each area of interest by the overall time spent attending towards any portion of the face. The findings revealed age differences in looking behaviour towards the eyes and mouth, which varied based upon the language participants heard (see Figure 10).

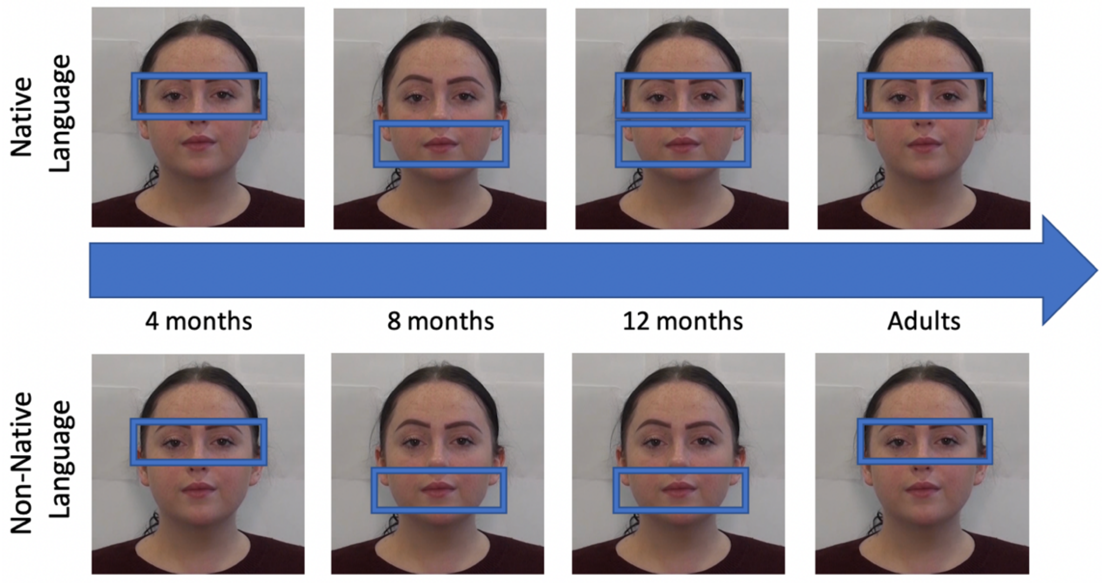


Figure 10: *Visual representation of face scanning behaviour at 4, 8, and 12 months, and in adults, towards a speaker producing native or non-native fluent speech as reported in Lewkowicz and Hansen-Tift (2012). Areas outlined in blue represent where on the speaker's face infants attended.*

Specifically, when viewing the speaker talk in their native language, Lewkowicz and Hansen-Tift (2012) found that 4-month-olds looked significantly longer at the eyes of the speaker rather than the mouth. However, at 6 months, infants began to display a shift in looking behaviour, looking equally towards the eyes and mouth. This attentional shift appeared to have fully occurred by 8 and 10 months, with infants at these ages focussing their attention on the speaker's mouth. Finally, 12-month-old infants showed another a shift in attention, this time looking equally at the speaker's mouth and eyes. The same pattern of results was found in infants viewing a non-native language being spoken, except at 12 months, when they preferentially looked at the mouth of the speaker. Adults were found to focus on the speaker's eyes irrespective of the language they heard. Taken together, this suggests that infants show two attentional shifts when viewing a face speak their native language during the first year of life, from the eyes to the mouth at 6 months, and then back towards an adult-like pattern of attending more to the eyes at 12 months.

Lewkowicz and Hansen-Tift (2012) proposed several explanations for the presence of attentional shifts in infants' audiovisual looking patterns. Firstly, visual attention develops rapidly throughout the first year of life (Colombo, 2001). In their first 6 months, infants' attention is primarily exogenously controlled, with visual fixations involuntarily driven by perceptual saliency. Given that the eyes are the most visually salient feature on a face (Shepherd et al., 1981), infants may be more likely to attend to a speaker's eyes at this age. However, infants begin to slowly develop endogenous attention from 6 months onwards. This allows them to inhibit salient stimuli and instead voluntarily control, as well as maintain, their visual fixations (Reynolds & Romano, 2016). For example, a face contained in a visual display with five distractors is preferentially attended to at 6 months, but not earlier at 3 months (Di Giorgio et al., 2012). Thus, the development of endogenous attention provides infants with the first opportunity to voluntarily gaze on either a speaker's eyes or mouth.

Secondly, an attentional switch from the eyes to the mouth occurs alongside the onset on canonical babbling (Oller, 2000). From 6 months onwards, infants start vocalising well-formed syllables comprised of a consonant-like and a vowel-like sound with a rapid transition between the two (e.g., /baba/). Lewkowicz and Hansen-Tift (2012) suggested that the increased motivation to produce speech sounds may drive an infant's attention towards the redundant audiovisual cues supplied by a speaker's mouth. Access to such highly informative signals may in turn assist speech and language acquisition. Indeed, attending to a speaker's mouth in the first year of life is linked to concurrent expressive language abilities (Tsang et al., 2018), and may also predict later vocabulary development (Tenenbaum et al., 2015; Young et al., 2009). Similarly, focussing on the mouth may help infants to learn how to produce their native speech sounds, with research finding that vocal imitations increase in 6-month-olds as a function of attending to a speaker's mouth (Imafuku et al., 2019).

Finally, the attentional shift found in 12-month-olds, back towards equally looking at a native speaker's eyes and mouth, may be linked to language experience. By the end of their first year, infants have acquired their native phonetic categories (Kuhl et al., 1992; Werker & Tees, 1984). This expertise may reduce the need for infants to access audiovisual cues when hearing their native language. This is further demonstrated by Lewkowicz and Hansen-Tift's (2012) finding that monolingual 12-month-olds continued to rely on the salient visual cues provided by the mouth when hearing an unfamiliar, non-native language. An additional study by Pons et al. (2015) further supports this claim. This study found that bilinguals, in comparison to monolinguals, looked longer at both a native and non-native speaker's mouth at both 4 and 12 months. Given that bilingual infants must learn different linguistic systems (Werker et al., 2009), an earlier mouth preference, which is then maintained, may assist with them in learning and differentiating between two languages. This again supports the idea that language experience can impact whether infants selectively maximise audiovisual cues at 12 months. The switch back towards the speaker's eyes can provide infants with additional assistance with their language acquisition. For example, following the gaze of a speaker can aid the formation of word-object associations in 12- to 18-month-olds (Barry-Anwar et al., 2017). Similarly, gaze following abilities during the first year of life are correlated with infants' receptive vocabulary at 14 and 18 months (Brooks & Meltzoff, 2005) and expressive vocabulary at 18, 21, and 24 months (Morales et al., 1998). Furthermore, attending to a speaker's eyes may subsequently boost an infant's social development. Indeed, Pons et al. (2019) found that, whilst 12-month-olds overall attended equally to a native speaker's eyes and mouth, infants who focused more on the eyes had higher socio-communicative skills.

Research has typically supported Lewkowicz and Hansen-Tift's (2012) findings that monolingual infants' selective attention to a native speaker's face changes during the first year of life (for a recent systematic review, see Bastianello et al., 2022). This has demonstrated that, when hearing fluent native speech, 4- to 6- month-olds will spend longer focussing on the eyes of a speaker (e.g., Pons et al., 2015; but see Morin-Lessard et al., 2019, for evidence of equal attention to the eyes and mouth at this age). Infants will then focus on the mouth of a speaker between 6 and 10 months (e.g., Morin-Lessard et al., 2019; Pons et al., 2015; but see Sekiyama et al., 2021, for evidence of attention to the eyes at this age). Finally, infants will then redirect their attention back towards an adult-like pattern of attending to the eyes by focussing equally on the eyes and mouth of a speaker at 12 months (e.g., Morin-Lessard et al., 2019; Pons et al., 2015, 2019; but see Tenenbaum et al., 2015, for evidence of maintaining a mouth preference at this age).

4.7 Infants' face scanning behaviour when viewing syllables

In addition to examining the attentional biases towards a speaker's eyes and mouth when presented with continuous speech, several studies have explored how audiovisually presented syllables modulate infants' typical face scanning behaviour. For example, Tomalski et al. (2013) presented 6- to 7- and 8- to 9-month-old infants, the majority of whom heard one language (English) in their home environment, with a series of videos displaying a single face producing repeating congruent (audio /ba/, visual /ba/; audio /ga/, visual /ga/), and McGurk inducing incongruent fusible (audio /ba/, visual /ga/) and incongruent non-fusible (audio /ga/, visual /ba/) stimuli. Overall attention to the face revealed that younger infants looked significantly less towards non-fusible stimuli compared to both congruent and fusible audiovisual pairings. In contrast, older infants attended equally to both incongruent trial types and less to congruent pairings. Thus, 8- to 9-month-olds did not visually discriminate in overall looking times between fusible and non-fusible audiovisual pairings, whereas 6- to 7-month-olds did. In relation to face scanning behaviour, no age-related differences in attention towards the speaker's eyes were found for any trial type. Age-related increases in looking towards the mouth were found, but only in non-fusible trials, with no changes observed between age groups for either congruent or fusible trials. Specifically, in fusible trials, both age groups attended more to the mouth compared to when viewing congruent stimuli. In non-fusible trials, younger infants fixated less on the mouth compared to congruent stimuli, whereas older infants focussed more on the visual articulations.

The findings of Tomalski et al. (2013) demonstrate that 6- to 9-month-old infants will adjust their attention to a speaker's mouth when viewing certain incongruent audiovisual stimuli. They also suggest that congruent syllables do not lead to age-related increases in attention to the mouth. This contrasts with the developmental increase in mouth looking behaviour observed for continuous speech in Lewkowicz and Hansen-Tift (2012). Methodological differences between the two studies, such as presenting congruent and incongruent syllables compared to only congruent fluent passages of speech, may have influenced infants' looking behaviour. Furthermore, as noted in Mercure et al. (2022), Tomalski et al. (2013) did not separate the data obtained from monolingual and bilingual infants. Given that language experience impacts face scanning behaviour when observing fluent speech (Pons et al., 2015), the inclusion of bilinguals may have influenced the overall findings in relation to attention to the mouth when viewing syllables.

Two more recent experiments have specifically examined monolingual infants' face scanning patterns when viewing incongruent syllables. Mercure et al. (2019) presented English-learning 4- to 8-month-olds with the same congruent, fusible, and non-fusible audiovisual pairings as Tomalski et al. (2013). Age-related increases in attention to the speaker's mouth were displayed in all three conditions. This provides evidence that a developmental shift for increased looking at a speaker's mouth is found with congruent syllables as well as continuous speech. An age-related sensitivity to incongruent syllables was also observed. Infants aged between 4 and 6.5 months, did not display any differences in face scanning behaviour for congruent and incongruent stimuli. However, 6.5 to 8-month-old infants were found to further increase their attention to the mouth when audiovisual cues were incongruent rather than congruent. The older infants also had a non-significant tendency to look longer at the mouth in non-fusible than in fusible incongruent trials. Together, these results demonstrate that incongruent audio and visual stimuli, at least for 6.5- to 8-month--old infants, are likely to lead to an increased preference towards attending to the mouth of a speaker.

A further study by Mercure et al. (2022), again using the same audio and visual pairings, measured face scanning behaviour in older 7- to 10-month-old English-learning infants. They found that infants attended equally to the speaker's mouth during congruent and fusible pairings. However, during non-fusible trials, infants showed an overall increase in their attention to the speaker's eyes rather than the speaker's mouth. The authors suggest that this result may reflect a more mature face scanning behaviour, whereby infants detected that the non-fusible mouth movements were not informative to the audio signal and so instead directed their attention towards the social information provided by the eyes. Indeed, there is evidence that 6- to 9-month-olds who shift their attention away from unhelpful visual cues during incongruent audiovisual trials exhibit better language abilities between 14 and 16 months compared to infants who do not (Kushnerenko et al., 2013).

There is research demonstrating that perceptual attunement can also impact infants' scanning behaviour towards congruent and incongruent audiovisual syllables from an unfamiliar language. Danielson et al. (2017) presented Canadian English-learning 6-, 9-, and 11-month-old infants with non-native Hindi /d̪a:/ and /ɖa:/ consonants that were either audiovisually matched or mismatched. Both 6- and 9-month-olds, ages in which infants are undergoing perceptual attunement, detected the audiovisual congruence of these non-native phonemes. Specifically, they increased their attention to the speaker's mouth in response to the mismatched audiovisual pairings but not to the matched audiovisual pairings. In contrast, 11-month-olds, who had presumably acquired their native phonemes, did not appear to detect the audiovisual incongruence. Instead, the older infants looked equally at the mouth when viewing matched and mismatched non-native audiovisual pairings. Thus, the ability to detect the audiovisual congruence for non-native consonants appears to decline following perceptual narrowing towards the end of the first year of life.

A different pattern of results has been found for face scanning behaviour towards audiovisually incongruent vowels. Pejovic et al. (2020) measured eye and mouth preferences in monolingual (Spanish or Basque) 4.5- and 8-month-olds when they viewed a face producing /i/ or /u/, vowel categories found in both Spanish and Basque. The audio and visual signals were either matched (i.e., audio /i/, visual /i/; audio /u/, visual /u/) or mismatched (i.e., audio /i/, visual /u/; audio /u/, visual /i/). For audiovisually matching pairings, infants looked longer at the speaker's eyes at 4.5 months. This switched to longer looks towards the mouth at 8 months. Furthermore, 4.5-month-olds appeared to recognise the incongruency between the mismatched audio and visual signals. However, rather than eliciting increased attention towards the speaker's mouth, as observed when viewing audiovisually incongruent consonants at this age (Mercure et al., 2019), infants instead looked longer towards the speaker's eyes. In contrast, audiovisual mismatches did not change infants' attention at 8 months. This suggests that audiovisual consonant and vowel processing may differ during an infant's first 12 months.

To conclude, infants' face scanning behaviours when viewing syllables follows a similar developmental trajectory as when viewing fluent speech. Infants will attend to a speaker's eyes at 4.5 months (Pejovic et al., 2020), display a reduction in attention to the eyes and an increase in attention to the mouth between 6 and 8 months (Mercure et al., 2019) and then focus on the mouth between 7 and 10 months (Mercure et al., 2022). However, this pattern of looking changes when viewing audiovisually incongruent consonants. Infants increase their overall looking time towards the speaker's mouth between 6 and 8 months (Mercure et al., 2019; Tomalski et al., 2013), and transition to looking more at the eyes between 7 and 10 months (Mercure et al., 2022). Similarly, incongruent vowels lead to an increase in attention to the speaker's eyes at 4.5 months, but no difference at 8 months (Pejovic et al., 2020). Thus, there is evidence that infants will adjust their looking patterns to the face when the audio and visual signals of syllables do not match.

4.8 Summary and present research

As outlined in the present chapter, infants are sensitive to the temporal correspondence between speech sounds and their associated mouth movements, audiovisually matching phonemes (Kuhl & Meltzoff, 1982; MacKain et al., 1983; Patterson & Werker, 2003) and fluent speech (Dodd, 1979; Kubicek et al., 2014) from very early in development. This ability narrows to their native phonemes (Danielson et al., 2017; Pons et al., 2009) and language (Kubicek et al., 2014; Lewkowicz & Pons, 2013) by their first birthday. Furthermore, an infants' perception of auditory speech can be altered by the visual information provided to them (Burnham & Dodd, 2004; Rosenblum et al., 1997; Weatherhead & White, 2017), demonstrating that they can successfully integrate auditory and visual signals. As with adults (e.g., Kim et al., 2004; Reisberg et al., 1987), the increased phonetic details provided by viewing a speaker's mouth, as well as the temporal relationship between audio and visual speech signals, can be advantageous to infants' speech perception during the first year of life. Audiovisual speech can improve infants' abilities to learn unimodal distributed phonemic contrasts (Teinonen et al., 2008), and help them to differentiate between non-native phonemes following perceptual attunement

(Ter Schure et al., 2016; but see Danielson et al., 2017), segment speech (Tan et al., 2023), and assist them with detecting and discriminating speech in noise (Lalonde & Werner, 2019). Together, this research suggests that, by the age of 12 months, infants can not only match audio and visual speech signals but benefit from their concurrent presentation.

Turning back to the development of the consonant bias, the benefits of viewing a speaker's mouth raises the question of whether providing British English-learning infants with audiovisual stimuli, rather than auditory-only, may augment their ability to differentiate between phonetic mispronunciations. To recap, the experimental data reported thus far in the present thesis has shown that, in head-turn preference tasks, British English-learning 5- and 11- to 12-month-olds' word form recognition is equally impacted by consonants and vowels. These findings differ particularly from French-learning infants, who exhibit a vowel bias at 5 months (Bouchon et al., 2015) and a consonant bias by 11 months in word form recognition tasks (Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020). However, from an acoustic-phonetic perspective, the higher ratio of consonants to vowels, increased vowel reduction, diphthongisation, and consonant clusters found in English compared to French, may contribute to British English-learning infants having more difficulties in detecting changes to familiar word forms in contrast to their French-learning peers. Therefore, it can be hypothesised that the additional phonetic and temporal properties present in audiovisual speech may boost British English-learning infants' phonetic discrimination of consonant mispronounced words and vowel mispronounced words. This increased phonetic perception may in turn impact whether infants perceive vowel mispronunciations as more comparable to correctly pronounced words than consonant mispronunciations.

The research presented in this chapter has also demonstrated that infants' face scanning behaviours towards a native speaker has two developmental shifts during their first year of life (Lewkowicz & Hansen-Tift, 2012). The first switch occurs at around 6 months, as infants transition from preferentially attending to a speaker's eyes to focussing more on the speaker's mouth. The second switch occurs by the age of 12 months, when infants will typically look equally towards a speaker's eyes and mouth when viewing a face producing fluent speech. However, at 12 months, infants will look longer at a speaker's mouth when it articulates a non-native language (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015). When hearing incongruent audio and visual pairings of consonants, infants will increase their attention to the mouth at around 6 to 8 months (Mercure et al., 2019; Tomalski et al., 2013), and towards the eyes at around 7 to 10 months (Mercure et al., 2022). For incongruent vowels, 4.5-month-olds look longer at the speaker's eyes whereas 8-month-olds do not alter their attention to the face (Pejovic et al., 2020). Thus, the typical face scanning pattern found in infants differs when they are presented with unfamiliar or incongruent stimuli.

In relation to the consonant bias, these findings raise the question of whether face scanning patterns will also differ when viewing a talking face produce mispronunciations of familiar word forms. Specifically, at 12 months, it can be expected that infants would divide their attention equally between the eyes and the mouth of a speaker pronouncing a familiar word, given their looking patterns for fluent speech (e.g., Lewkowicz & Hansen-Tift, 2012). However, if infants detect mispronunciations, then it can be predicted this typical attention to the face would change, given that unfamiliar or unexpected audio and visual pairings lead to an increase in looking towards a speaker’s eyes or mouth (Lewkowicz & Hansen-Tift, 2012; Mercure et al., 2019, 2022; Pejovic et al., 2020; Pons et al., 2015; Tomalski et al., 2013). Furthermore, the change in the pattern of looking may differ based upon whether the mispronunciation is made to the word’s onset consonant or vowel.

The impact of audiovisually presented stimuli on the presence of a consonant bias in British English-learning infants will be explored in two eye-tracking experiments in Chapter Four. These studies will examine whether 12-month-old British English-learning infants display a looking preference towards a face articulating lists of familiar word forms (Experiment 4) or a single word form (Experiment 5) when they are spoken either correctly, with an onset consonant mispronunciation, or with a vowel mispronunciation. Both studies will also investigate whether infants’ face scanning behaviours towards the speaker’s eyes and/or mouth differ between correctly pronounced and mispronounced words.

5 Chapter Four

5.1 Experiment 4: Phonetic specificity of audiovisually presented familiar word forms in British English-learning 12-month-olds

5.1.1 Introduction

Experiment 4 continued the investigation into the phonetic specificity of British English-learning 12-month-olds' familiar word form recognition. However, rather than using an auditory head-turn preference procedure to present infants with phonetic changes to familiar word forms (as in the research presented in Chapter Two), this study examined whether audiovisually presented stimuli impacted the presence of the consonant bias. Recall that, by 12 months, infants are attuned to their native vowel (e.g., Polka & Werker, 1994) and consonant (e.g., Werker & Tees, 1984) categories. Similarly, infants' early ability to match audio and visual syllables and speech (e.g., Dodd, 1979; Kuhl & Meltzoff, 1982) has narrowed to their native phonemes and language by their first birthday (Danielson et al., 2017; Kubicek et al., 2014; Pons et al., 2009). Infants can also differentiate between familiar word forms and pseudowords, both when presented auditorily (e.g., Hallé & de Boysson-Bardies, 1994; Vihman et al., 2004) and audiovisually (Weatherhead & White, 2017). Thus, 12 months is an ideal age to explore whether audiovisual speech impacts infants' phonetic discrimination.

The present experiment had two main aims. Firstly, it examined the modulating effects of audiovisual information on the reliance of consonants over vowels in British English-learning infants' word form recognition. To investigate this, the study adopted a comparable methodology used in research investigating infants' face scanning patterns and preferences when presented with videos of audiovisually mismatched syllables (e.g., Mercure et al., 2019, 2021; Pejovic et al., 2020; Tomalski et al., 2013). Specifically, infants were shown videos of a face articulating lists of familiar word forms that were either correctly pronounced or mispronounced with a one-feature phonetic change to either the word's initial consonant or vowel. To maintain consistency with the audio-only research presented in Chapter Two, the familiar words were identical to the stimuli infants displayed a familiarity preference for when contrasted with pseudowords in Experiment 2a. Similarly, the consonant and vowel mispronunciations were the same as the mispronunciations that infants did not auditorily discriminate between in Experiment 2b. Recall that these words contained several types of feature changes made to the onset consonants (4 place, 3 voice, 3 manner) and vowels (7 backness, 3 height), which differ in their visual distinctiveness.

Using an eye-tracker, the total time infants spent fixated on the speaker's entire face was recorded, along with their attention to the speaker's eyes and mouth. Based upon previous research demonstrating that British English-learning infants do not display a consonant bias in lexical processing tasks until approximately 30 months (Delle Luche et al., 2017; Floccia et al., 2014; Mani & Plunkett, 2010; Nazzi et al., 2009), it could be expected that infants here would again not differentiate between correct pronunciations of familiar word forms and either the consonant or vowel mispronunciations. Alternatively, given that the additional phonetic detail provided by audiovisual speech has been found to positively impact infants' speech perception (e.g., Teinonen et al., 2008; Ter Schure et al., 2016), the addition of visual speech cues may be sufficient to boost British English-learning infants' phonetic perception and consequently enhance their ability to detect the consonant and vowel alterations made to familiar word forms. Together, this may in turn lead them to discriminate between pronunciation types and display a consonant bias in word form recognition, as found in their French-learning peers at this age (Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020). Furthermore, infants may show a preference for the unfamiliar sounding stimuli (i.e., the consonant mispronunciations), given past research demonstrating that infants attend longer to audiovisually presented unfamiliar words over familiar words (Weatherhead & White, 2017).

The second objective of this experiment was to assess whether infants' attention to specific areas of the face differed for each pronunciation type. For correct pronunciations, it was predicted that infants would display similar face scanning behaviours as previously found when observing fluent speech in their native language at 12 months (e.g., Lewkowicz & Hansen-Tift, 2012), dividing their attention equally between the speaker's eyes and mouth. Similarly, if British English-learning 12-month-olds treated consonant and vowel mispronunciations of familiar word forms comparably to their correct pronunciations, then they could be expected to maintain the same pattern of attention when observing such alterations.

However, whilst no research has directly examined infants' gaze behaviours for mispronounced words, several past findings can lead to the hypothesis that attention towards a speaker's eyes and mouth may alter when hearing mispronunciations. Specifically, infants increase their attention to the speaker's mouth at 12 months when it is articulating a non-native language (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015) and between 6 and 8 months when the audio and visual signals of consonants are mismatched compared to when they are matched (Mercure et al., 2019; Tomalski et al., 2013). Conversely, infants have been found to increase their attention to the speaker's eyes at 4.5 months, but not at 8 months, for auditorily and visually incongruent vowels (Pejovic et al., 2020) and between 7 and 10 months for incongruent consonants (Mercure et al., 2022). Thus, if British English-learning infants detect the mispronunciations to familiar word forms, they may modify their fixations towards the face, through increased attention to either the eyes or the mouth of the speaker. Furthermore, such variations in looking behaviour may differ based upon the phonetic alteration presented to infants.

In addition, the present study measured infants' vocabulary size and calculated the number of unique consonant and vowel tiers they knew. This was again done to investigate the possible relationship between lexical development and the presence of the consonant bias, as well as exploring whether lexical factors impacted infants' face scanning behaviour.

5.1.2 Method

Participants

A total of 25 healthy British English-learning monolingual 12-month-old infants were successfully tested (mean age = 11;20, range = 10;27 to 12;22, 14 females, 11 males). The data of an additional nine infants were excluded from the analysis due to crying or being distracted (n = 7) or parental interference (n = 2). All parents completed the 100-word Oxford Short Form CDI (Floccia et al., 2018; Hamilton et al., 2000) and a checklist of the 10 test words presented in the study as a measure of each infant's vocabulary. All infants were raised in the South West of England.

Materials

The stimuli were videos of a native female British English speaker from the South West of England, filmed from the shoulders up against a plain, light background. All videos were recorded in high definition using a digital camcorder (Sony HXR-NX30E) with a top-mounted microphone and were edited using Adobe Premiere Pro (Adobe Systems, San Jose, CA, USA). One video for each of the 10 correct pronunciations of words heard in Experiment 2a (refer back to Table 4), and the 10 consonant mispronunciations and 10 vowel mispronunciations heard in Experiment 2b (refer back to Table 6) was recorded, resulting in a total of 30 videos.

All videos were produced by the speaker in an infant directed voice whilst maintaining a neutral expression and looking directly at the camera. The videos were all 2 seconds in length and were edited to ensure that the speaker began and finished with their mouth closed and eyes open to allow for videos to be played consecutively with minimal visual disruption. Praat was used to clean the audio files of each video to reduce any background noise and normalise the amplitude at 70dB. The acoustic features of the stimuli used in each condition are listed in Table 10. No significant statistical differences in duration or mean, minimum, or maximum fundamental frequency between the audio in all videos.

Table 10: *Acoustic features of stimuli presented in Experiment 4.*

	Condition			t-value and significance level (two-sided)
	Correct	Consonant Change	Vowel Change	
Duration (ms)	749 (38)	747 (52)	742 (52)	$t(9) = .14$, $p = .89$, $t(9) = .40$, $p = .70$, $t(9) = .26$, $p = .80$
F0 mean (Hz)	233 (17)	234 (14)	232(16)	$t(9) = -.17$, $p = .87$, $t(9) = .60$, $p = .56$, $t(9) = .98$, $p = .35$
F0 min (Hz)	189 (33)	188 (34)	187 (39)	$t(9) = .15$, $p = .89$, $t(9) = .39$, $p = .71$, $t(9) = .13$, $p = .90$
F0 max (Hz)	285 (21)	288 (15)	288 (10)	$t(9) = -.38$, $p = .71$, $t(9) = -.48$, $p = .65.$, $t(9) = .07$, $p = .95$

Note. Standard deviations are reported in parentheses. The first *t*-test on each line corresponds to the comparison between correct and consonant change, the second to the comparison between correct and vowel change, and the third the comparison between consonant change and vowel change.

The 10 videos in each condition (correct pronunciations, consonant mispronunciations, vowel mispronunciations) were concatenated into pseudo-randomised lists with each video appearing once. The position that each video was presented in each list was evenly distributed across lists. This resulted in lists that were 20 seconds in length. The videos measured 792 x 672 pixels and were presented in the centre of a 1080 x 1920 pixels screen.

Procedure

Upon obtaining informed consent from the parent, infants were strapped into a high-chair facing a screen in a sound-attenuated and dimly lit testing booth (see Appendix B for a diagram of the experimental setup). Infants were positioned approximately 60cm away from the screen. The parent stood behind the infant, out of the range of the eye-tracker, and were asked not to influence their infant's attention or point towards the screen. Infants' eye movements were recorded using a Tobii 300 TX, with a sampling rate of 120 Hz, which was mounted underneath the screen. The auditory stimuli were presented through loudspeakers located 15cm either side of the visual display. The experimenter initiated the trials and observed the infant's eye movements on a separate display outside of the booth.

The eye-tracker was first calibrated to the infant's eyes using a standard five-point calibration routine (the four corners and centre of the screen) to obtain the infant's point-of-gaze. During calibration, a star shape would move sequentially to each of the five target points. Once calibration was completed, using a within-participant design, infants then saw a total of 12 test videos, consisting of four lists for each condition (correct pronunciations, consonant mispronunciations, vowel mispronunciations). Trials were split into two blocks, with each block consisting of two lists of each condition. The presentation of each condition in each block was randomised. Each trial was proceeded by an attention getter, a smiling cartoon face, displayed centrally on the screen. The experimenter would begin the trial when the infant's gaze was in the centre of the screen. During each trial, the video lists of the speaker would play in the centre of the screen for 20 seconds. The video remained on the screen for the duration of the trial, after which the process would repeat. The study lasted for approximately five minutes.

5.1.3 Results and discussion

To examine infants' looking preferences, fixations towards three areas of interest (AOI) were extracted using Tobii Studio. As depicted in Figure 11, the AOIs corresponded to the area encompassing the speaker's eyes (140 x 400 pixels), the area encompassing the speaker's mouth (120 x 240 pixels), and the area covering the speaker's face (600 x 500 pixels). The size of the AOIs were selected to be comparable to past research examining infants' attention to a speaker's eyes and mouth (e.g., see Ter Schure et al., 2016) The three areas remained fixed for the duration of the experiment and were large enough to surround the relevant AOI in each video, allowing for any small movements made by the speaker.

Figure 11: *Example screenshot of the videos presented in Experiments 4 and 5 with the three areas of interest (AOI) added for reference. The eye AOI is coloured in yellow, the mouth AOI is coloured in pink, and the face AOI is outlined in red.*

Prior to analysing the data, trials were removed based upon exclusion criteria used in previous research with a comparable methodology (Mercure et al., 2019, 2022). This excluded any trials from the analysis when infants attended to the entire face for less than a total of 3 seconds and only included infants with at least 70% of successful trials. This resulted in 10 out of the 300 trials (3.33%) being removed from the final analysis. All included infants had at least 10 successful trials out of the possible 12 (83.33%).

Overall looking time to the face

The total looking times to all three AOIs (eyes, mouth, face) were calculated for each trial to provide an overall looking time for each pronunciation type. The mean looking times for the correct (*M* = 12.34, *SD* = 3.29), consonant (*M* = 12.65, *SD* = 2.57), and vowel (*M* = 12.97, *SD* = 3.59) conditions are displayed in Figure 12.

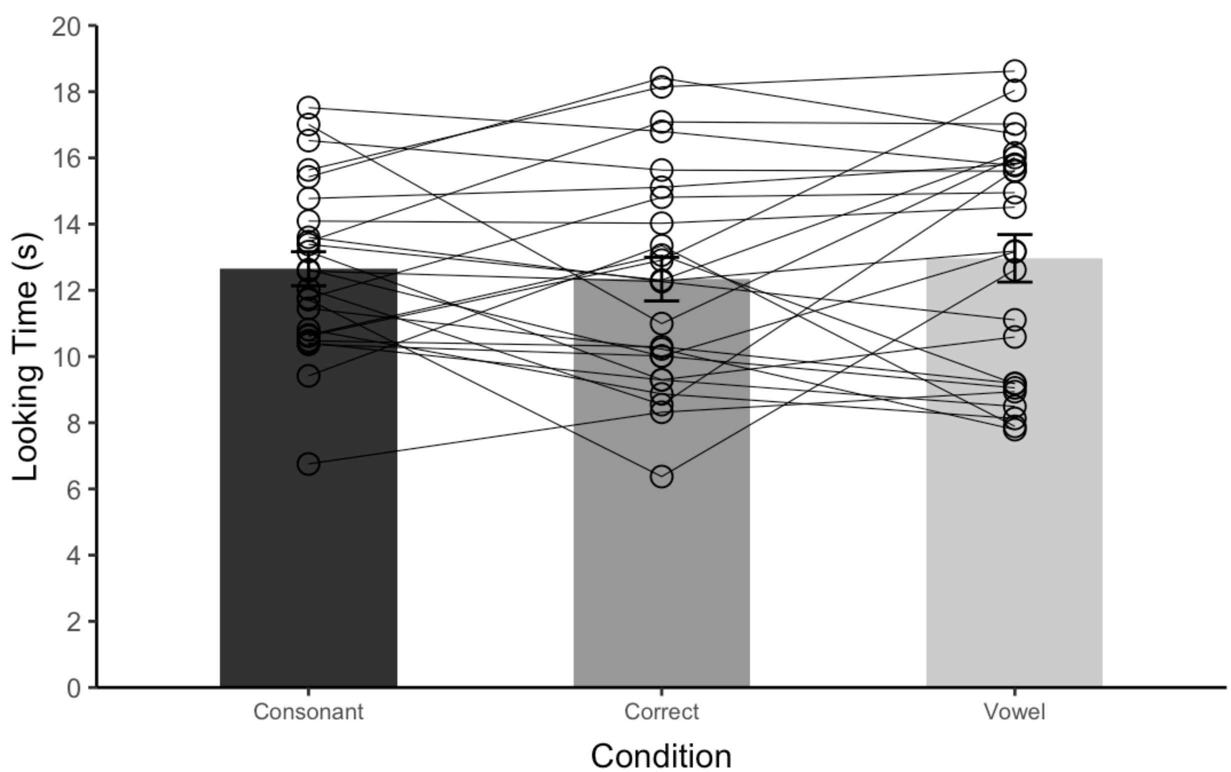


Figure 12: *Mean looking times (s) to the speaker's face in each of the three experimental conditions In Experiment 4. Error bars represent ±1 standard error.*

Shapiro Wilk tests for normality found that looking times towards the correct pronunciations (W = .97, p = .51) and the consonant changes (W = .97, p = .74) were normally distributed but looking times towards the vowel changes were not (W = .90, p = .02). Consequently, raw looking times were log-transformed before analysis (Csibra et al., 2016).

A repeated measures ANOVA on log-transformed overall looking times, with Pronunciation (correct, consonant, vowel) and Block (first half versus second half of the experiment) as within-participant factors, found a significant effect of Block, $F(1, 24) = 45.77$, $p < .001$, $\eta_G^2 = .22$, with infants attending to the speaker more during the first half of the study. However neither the effect of Pronunciation, $F(2, 48) = .58$, $p = .56$, $\eta_G^2 < .001$, nor the Pronunciation x Block interaction, $F(2, 48) = .42$, $p = .66$, $\eta_G^2 = .002$, reached significance. Additionally, the ANOVA was rerun without the effect of Block, again revealing no significant effects of Pronunciation, $F(2, 48) = .52$, $p = .60$, $\eta_G^2 = .01$.

The reliability of the null finding (no effect of condition) was also examined with paired Bayesian *t*-tests, comparing log-transformed looking times for each pronunciation type. The comparison of correct versus consonant looking times had a Bayes Factor of BF = .29 ($t(24) = -.84$, $p = .41$), the comparison of consonant versus vowel overall looking time had a Bayes Factor of BF = .21 ($t(24) = -.13$, $p = .90$), and the comparison of correct versus vowel overall looking time found a Bayes Factor of BF = .29 ($t(24) = -.83$, $p = .41$). Thus, all comparisons indicated support for the null hypothesis.

Effect of mispronunciation on face scanning behaviour

The amount of time infants focussed on the speaker's eyes and mouth was determined by using the same calculations as previous research examining infants' face scanning behaviour (e.g., Lewkowicz & Hansen-Tift, 2012; Mercure et al., 2019, 2022; Pons et al., 2015). Specifically, the proportion-of-total-looking-time (PTLT) that infants spent attending to the eyes and mouth was calculated by dividing the total time spent focussing on each AOI by the total amount of time spent focussing on any area of the face. The mean Eye-PTLT and Mouth-PTLT scores, as well as mean looking times to each AOI, are listed in Table 11.

Table 11: *Mean looking times (s) to the three areas of interest on the speaker's face, as well as eyes-to-face and mouth-to-face ratios, for correct pronunciations, consonant mispronunciations, and vowel mispronunciations in Experiment 4.*

Pronunciation	Eyes	Mouth	Face	Eyes-to-Face	Mouth-to-Face
Correct	4.02 (3.14)	5.64 (3.06)	2.68 (2.68)	.32 (.20)	.44 (.19)
Consonant	4.36 (2.61)	5.63 (2.96)	2.67 (1.73)	.35 (.20)	.43 (.19)
Vowel	3.93 (2.64)	6.23 (3.45)	2.80 (1.62)	.31 (.18)	.46 (.18)

Note. Standard deviations are reported in parentheses.

Additionally, PTLT difference scores were calculated for every trial by subtracting each infant's eye-PTLT scores from their mouth-PTLT scores. The mean PTLT difference scores for the consonant (M = -.09, SD = .36), correct (M = -.11, SD = .36), and vowel (M = -.16, SD = .34) conditions are plotted in Figure 13. Shapiro-Wilk tests for normality found that infants' PTLT difference scores for the correct pronunciation (W = .99, p = .98), the consonant change (W = .98, p = .80), and the vowel change (W = .95, p = .22) were all normally distributed.

A repeated measures ANOVA on PTLT scores with Pronunciation (correct, consonant, vowel), AOI (eyes, mouth), and Block (first half versus second half of the experiment) as within-participant factors found that the main effects of Pronunciation, $F(2, 48) = .34$, $p = .72$, $\eta_G^2 = 2.12$, AOI, $F(1, 24) = 2.99$, $p = .10$, $\eta_G^2 = .08$, and Block, $F(1, 24) = .61$, $p = .44$, $\eta_G^2 = 2.78$, were all not significant. Similarly, no significant interactions were observed ($p > .05$ in all cases). The ANOVA was rerun without the effect of Block, again finding no significant effect of Pronunciation ($F(2, 48) = .33$, $p = .72$, $\eta_G^2 = 2.38$), or AOI ($F(1, 24) = 3.08$, $p = .09$, $\eta_G^2 = .09$), nor a Pronunciation x AOI interaction ($F(2, 48) = 2.19$, $p = .12$, $\eta_G^2 = .01$).

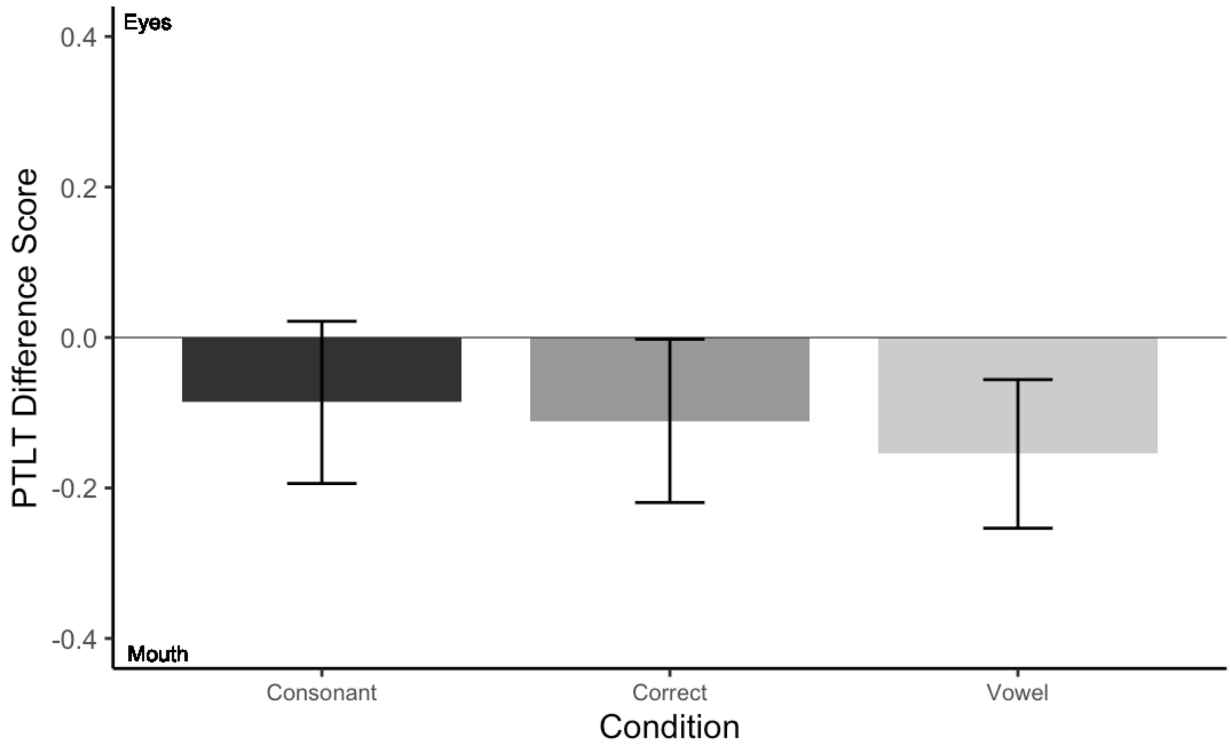


Figure 13: *Mean proportion-of-total-looking-time (PTLT) scores across the three conditions in Experiment 4. Positive values indicate a preference for the eye region of the face. Negative values indicate a preference for the mouth region of the face. Error bars represent ±1 standard error.*

The reliability of this null finding was again further explored with paired Bayesian *t*-tests, comparing the mean PTLT scores for each pronunciation type. These found that correct versus consonant PTLT scores had a Bayes Factor of BF = .30 ($t(24) = -.88$, $p = .38$), indicating support for the strength of the null hypothesis. In contrast, correct versus vowel PTLT scores had a Bayes Factor of BF = .35 ($t(24) = 1.07$, $p = .29$) and consonant versus vowel PTLT scores had a Bayes Factor of BF = 2.24 ($t(24) = 2.39$, $p = .03$), which provided inconclusive evidence for either the alternative or null hypothesis.

Vocabulary measures and face looking behaviour

To correct for multiple comparisons, a Bonferroni correction was applied to allow for 16 comparisons (.05/16 = .003). Infants had an average CDI comprehension of 9.56 words (*SD* = 9.32) out of 100 and knew an average of 5.04 (*SD* = 2.39) of the 10 words audiovisually presented in the experiment. No correlations were found between infants' log-transformed listening preferences, (correct versus consonant, correct versus vowel, consonant versus vowel) and their CDI or word list comprehension ($p > .003$ in all cases, BF = .25 to .60). Infants knew an average of 13.24 (*SD* = 10.98) of the possible 108 unique words found by combining the Oxford Short Form CDI and the study checklist, from which there was an average 12.40 (*SD* = 9.85) consonant tiers and 10.36 (*SD* = 6.36) vowel tiers. The proportion of unique consonant tiers from the known words (*M* = .96, *SD* = .05) was found to be significantly different to the proportion of vowel tiers (*M* = .89, *SD* = .14), $t(24) = 3.09$. $p = .01$. However, no correlation was found between infants' log-transformed listening preferences and either the proportion of consonant or vowel tiers they knew ($p > .003$ in all cases, BF = .26 to .48). Finally, no correlation was found between infants' overall PTLT score (*M* = -.12, *SD* = .34) and either CDI and word list comprehension, or the number of consonant and vowel tiers they were familiar with ($p > .003$ in all cases, BF = .27 to .49). All correlations are listed in Appendix C.

The results of Experiment 4 provide additional evidence that 12-month-old British English-learning infants do not differentiate between consonant and vowel mispronunciations made to lists of familiar word forms. In relation to an overall preference for each pronunciation type, infants attended equally to each audiovisual list, irrespective of whether the words were correctly pronounced, or mispronounced with an alteration to the word's initial consonant or vowel. This indicates that, even when providing the additional phonetic cues afforded by viewing a speaker's mouth, infants did not demonstrate a consonant bias in word form recognition. This replicates the findings of Experiment 2b, in which infants did not distinguish between the same consonant and vowel mispronunciations in an auditory-only head-turn preference study. Consequently, this result adds to the growing evidence that British English-learning infants do not display a consonant bias in lexical tasks until approximately 30 months (Floccia et al., 2014; Nazzi et al., 2016).

Furthermore, evidence of infants' face scanning behaviour suggests that they did not adjust their attention to the speaker's eyes or mouth when hearing the mispronunciations of familiar word forms. For the correct pronunciation, no significant differences were found for where on the face infants attended. This finding replicated face scanning patterns found in 12-month-olds when they are presented with passages of continuous speech, with infants looking equally to the eyes and mouth of a talking face (Lewkowicz & Hansen-Tift, 2012). In relation to attention to the face when viewing audiovisual mispronunciations, past research has shown that infants at 12 months will focus more on a speaker's mouth when hearing an unfamiliar language (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015) and adjust their typical face scanning behaviour towards a speaker's face during the first year of life when viewing incongruent audio and visual syllables (Mercure et al., 2019, 2021; Pejovic et al., 2020; Tomalski et al., 2013). However, when presented with the phonetic changes to the familiar word forms, infants did not adjust their attention to a speaker's eyes or mouth for either consonant or vowel alterations. Thus, such mispronunciations appeared to be visually attended to in a comparable way to correctly pronounced words.

Finally, CDI scores and the understanding of the words presented to infants, as well as the proportion of consonant and vowel tiers from these words, did not predict looking infants' preferences. Therefore, lexical factors did not appear to impact whether infants display a consonant bias in the audiovisual task presented in this experiment.

5.2 Experiment 5: Phonetic specificity of an audiovisually presented single familiar word form in British English-learning 12-month-olds

5.2.1 Introduction

In the final experiment of the thesis, British English-learning 12-month-olds' preference towards videos of a speaker, as well as their face scanning behaviour, was again examined in an eye-tracking study. However, rather than viewing a speaker produce lists of familiar word forms, as in Experiment 4, infants saw the speaker articulate a single word form. Specifically, the word presented to infants was *mummy*, pronounced correctly, with an onset consonant mispronunciation (*nummy*; /nʌmi/), and with a vowel mispronunciation (*memmy*; /mɛmi/). Infants' attention towards the correct pronunciation of the word *mummy* versus either a consonant or vowel mispronunciation was previously examined in a between-subjects design using the head-turn preference paradigm in Experiment 3. The results of this study found an overall mispronunciation effect, with infants attending longer to the correct pronunciation of *mummy* compared to either a consonant alteration or a vowel alteration. However, this preference was comparable across both consonant and vowel mispronunciation conditions. Therefore, infants did not appear to show a consonant bias in this audio-only word recognition task.

Nevertheless, there are several factors which may lead infants to display different looking patterns when presented with a speaker visually repeating a single word compared to lists of words. Viewing the articulations of a single word is more comparable to research which has found that infants' attention to the eyes and mouth of a speaker differ when presented with repetitions of a single congruent or incongruent audiovisual syllable (e.g., Mercure et al., 2019, 2022). Observing a speaker articulate a single repeated word, as opposed to a randomised list of words, may help to reduce infants' working memory load during the task. Finally, the consonant mispronunciation was likely to be highly salient to infants, given that the feature change was made to the place of articulation (Ladefoged & Disner, 2012). This contrasted with the variations in visual distinctiveness of the consonant alterations presented in Experiment 4 (4 place, 3 voice, 3 manner). Indeed, there is evidence to suggest that voiced changes are treated comparably to their correct pronunciations by infants at the end of their first year (Weatherhead & White, 2017). Together, these factors could increase the likelihood that a consonant bias would be observed.

5.2.2 Method

Participants

A further 25 healthy British English-learning monolingual 12-month-old infants raised in the South West of England were successfully tested (mean age = 11;27, range = 11;9 to 12;22, 14 females, 11 males). The data of an additional 13 infants were excluded from the analysis due to crying or being distracted (*n* = 4), parental interference (*n* = 2), being unable to achieve a successful calibration (*n* = 4), or not meeting the looking time criteria (*n* = 3). The 100-word Oxford Short Form CDI (Floccia et al., 2018; Hamilton et al., 2000) was again completed by 24 of the 25 parents as a measure of each infant's vocabulary.

Materials

The audiovisual stimuli were recorded using the same speaker and setup as in Experiment 4. The videos consisted of the speaker staring directly into the camera and saying the target word (*mummy*; /mʌmi/), correctly pronounced, with a one-feature change to the initial consonant (*nummy*; /nʌmɪ/), and with a one-feature change to the initial vowel (*memmy*; /memɪ/). To help maintain infants' interest to the stimuli, four different videos of each pronunciation were recorded, each with a slightly different intonation. Intonations were matched across pronunciation types. The audio of each video was again cleaned, and the amplitude normalised to 70dB using Praat. Each video was 2 seconds in length and was played ten times in each trial. Thus, each trial was 20 seconds in length.

The acoustic features of the stimuli used in each condition are listed in Table 12. There were no significant statistical differences in duration or mean, minimum, or maximum fundamental frequency between the audio in all videos.

Table 12: *Acoustic features of stimuli presented in Experiment 5.*

	Condition			t-value and significance level (two-sided)
	Correct	Consonant Change	Vowel Change	
Duration (ms)	791 (127)	784 (80)	807 (95)	$t(6) = .09$, $p = .93$, $t(6) = -.20$, $p = .85$, $t(6) = -.37$, $p = .72$
F0 mean (Hz)	318 (30)	319 (39)	329 (42)	$t(6) = -.03$, $p = .97$, $t(6) = -.42$, $p = .69$, $t(6) = -.03$, $p = .74$
F0 min (Hz)	274 (49)	283 (55)	294 (60)	$t(6) = -.25$, $p = .81$, $t(6) = -.52$, $p = .62$, $t(6) = -.27$, $p = .79$
F0 max (Hz)	390 (33)	382 (21)	394 (18)	$t(6) = .37$, $p = .72$, $t(6) = -.24$, $p = .82$, $t(6) = -.85$, $p = .43$

Note. Standard deviations are reported in parentheses. The first *t*-test on each line corresponds to the comparison between mummy and nummy, the second to the comparison between mummy and memmy, and the third the comparison between nummy and memmy.

Procedure

The study used the same experimental procedure and apparatus as detailed in Experiment 4. Infants were presented with a total of 12 trials, comprised of four for each condition (correct pronunciation, consonant mispronunciation and vowel mispronunciation). The video sequence remained on the screen for the duration of each trial. The trials were again organised into two blocks. Each block contained two presentations of each pronunciation type. The order of pronunciation types within each block were randomised.

5.2.3 Results and Discussion

Fixations towards the three AOIs were extracted from Tobii Studio. The AOIs had the same dimensions as in Experiment 4. All trials below 3 seconds were again removed from the analysis. This resulted in 10 out of the 300 trials (3.33%) being removed from the final analysis. Each included infant had at least 10 successful trials out of the possible 12 (83.33%) presented. Shapiro-Wilk tests for normality revealed that infants' looking times to the correct pronunciation (W = .93, p = .11) and the consonant change (W = .94, p = .17) were normally distributed but looking times to the vowel change were not normally distributed (W = .90, p = .02).

Overall looking time to the face

As in Experiment 4, the total looking times to the three different AOIs (eyes, mouth, face) were calculated for each trial. Figure 14 displays the mean looking times in the consonant (*M* = 11.43. *SD* = 2.86), correct (*M* = 12.21, *SD* = 3.00), and vowel (*M* = 12.64 *SD* = 2.97) conditions.

A repeated measures ANOVA on overall log-transformed looking times with Pronunciation (correct, consonant, vowel) and Block (first half versus second half of the experiment) as within-participant factors found a significant effect of Pronunciation, $F(2, 48) = 5.27$, $p = .01$, $\eta_G^2 = .02$, a significant effect of Block, $F(2, 46) = 84.40$, $p < .001$, $\eta_G^2 = .24$, but a non-significant Pronunciation x Block interaction, $F(2, 48) = .37$, $p = .69$, $\eta_G^2 = .69$. Given that Block did not interact with Pronunciation, the statistical analysis was rerun excluding Block from the ANOVA. This again found a significant effect of Pronunciation, $F(2, 46) = 5.84$, $p = .01$, $\eta_G^2 = .02$.

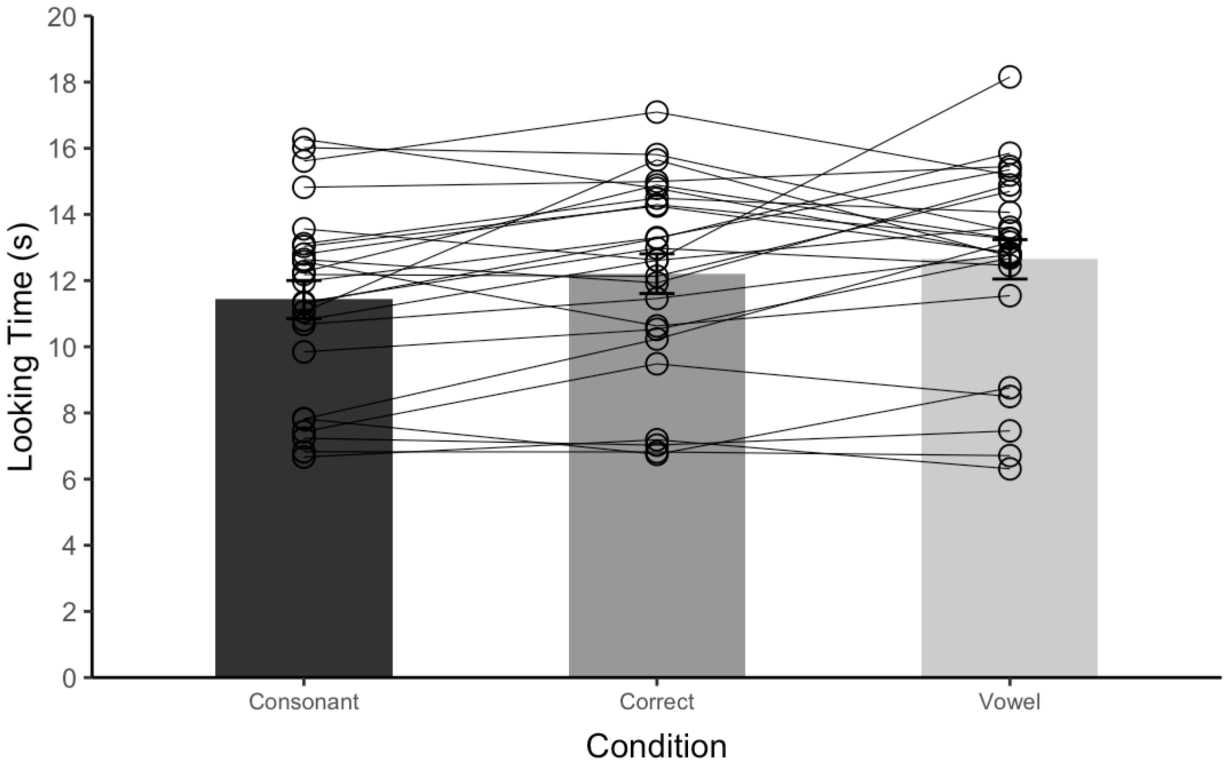


Figure 14: *Mean looking times (s) to the speaker's face in each of the three experimental conditions in Experiment 5. Error bars represent ±1 standard error.*

Three paired Bayesian *t*-tests were computed to compare each pronunciation type. Log- transformed looking times for correct versus consonant had a Bayes factor of BF = 3.19 ($t(24) = 2.59$, $p = .02$), which provided evidence for the strength of the alternative hypothesis. Similarly, log-transformed looking times for consonant versus vowel had a Bayes factor of BF = 8.27 ($t(24) = -3.07$, $p = .01$), which also provided evidence for the strength of the alternative hypothesis. Log-transformed looking times for correct versus vowel had a Bayes factor of BF = .38 ($t(24) = -1.15$, $p = .26$), which indicated inconclusive evidence for either the alternative or null hypothesis.

Effect of mispronunciation on face scanning behaviour

The proportion of looking time to the speaker's eyes and mouth, as well as PTLT difference scores, were calculated in an identical manner to Experiment 4. The mean Eye-PTLT and Mouth-PTLT scores and mean looking times to each AOI, are listed in Table 13.

Table 13: *Mean looking times (s) to the three areas of interest on the speaker's face, as well as eyes-to-face and mouth-to-face ratios, for correct pronunciations, consonant mispronunciations, and vowel mispronunciations in Experiment 5.*

Pronunciation	Eyes	Mouth	Face	Eyes-to-Face	Mouth-to-Face
Correct	3.64 (2.62)	6.29 (3.53)	2.29 (.26)	.30 (.19)	.50 (.22)
Consonant	4.19 (3.08)	5.16 (3.09)	2.08 (.73)	.36 (.22)	.44 (.22)
Vowel	4.67 (3.42)	5.70 (3.30)	2.27 (.04)	.36 (.23)	.45 (.23)

Note. Standard deviations are reported in parentheses.

The mean PTLT difference scores for the consonant (*M* = -.08, *SD* = .43), correct (*M* = -.20, *SD* = .41), and vowel (*M* = -.09, *SD* = .45) conditions are plotted in Figure 15. Shapiro-Wilk tests for normality found that infants' PTLT difference scores for the correct pronunciation (W = .95, p = .31), the consonant change (W = .95, p = .24), and the vowel change (W = .93, p = .07) were all normally distributed.

A repeated measures ANOVA on PTLT scores with Pronunciation (correct, consonant, vowel), AOI (eyes, mouth), and Block (first half versus second half of the experiment) as within-participant factors found a significant effect of Block, $F(2, 48) = 15.41$, $p < .001$, $\eta_G^2 = .002$, a significant Block x AOI interaction, $F(2, 48) = 4.83$, $p = . 04$,

η_G^2 = .01 and a significant Pronunciation x AOI interaction, $F(2, 48) = 6.62$, $p < .01$, η_G^2 = .01. No other significant effects were found ($p > .05$ in all cases).

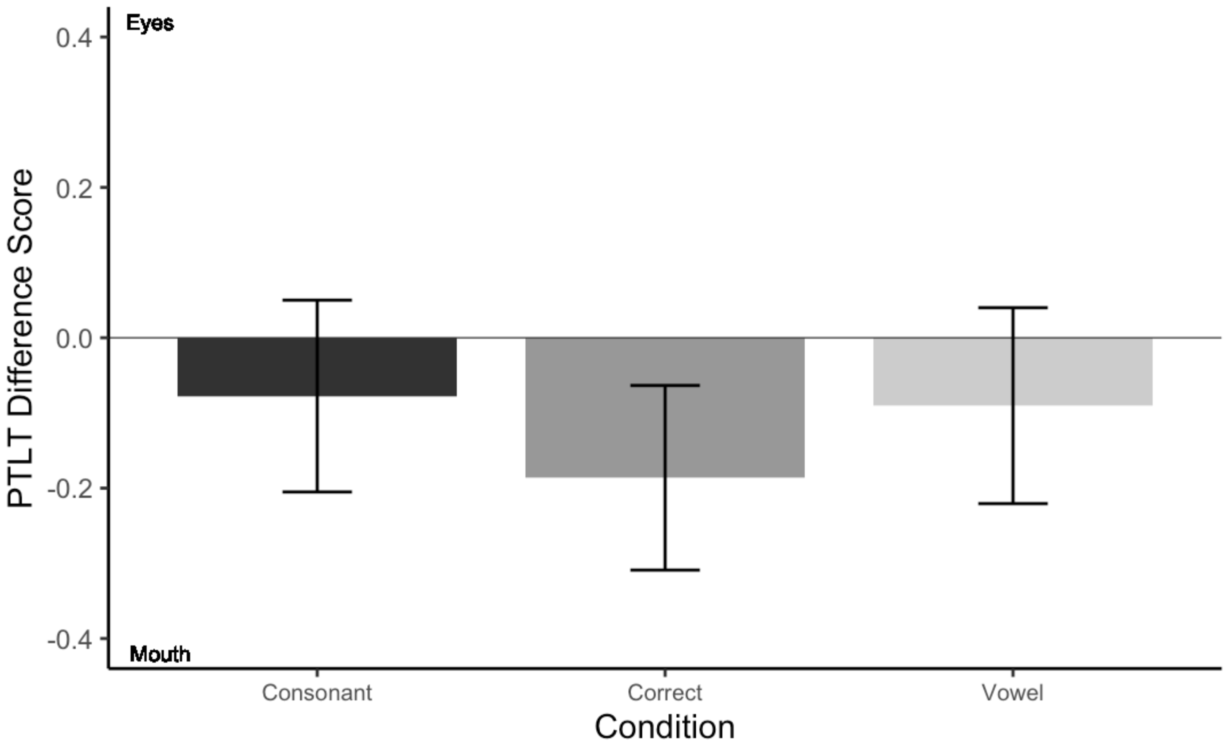


Figure 15: *PTLT scores across the three conditions in Experiment 5. Positive values indicate a preference for the eye region of the face. Negative values indicate a preference for the mouth region of the face. Error bars represent ±1 standard error.*

In relation to the Block x AOI interaction, paired Bayesian *t*-tests, comparing the mean looking time to each AOI in Block 1 and Block 2 were calculated. These found that the comparison of the PTLT towards the mouth in Block 1 (*M* = .49, *SD* = .24) and Block 2 (*M* = .43, *SD* = .21) had a Bayes Factor of BF = 5.57 ($t(24) = 2.88$, $p = .001$. In contrast, the comparison of the PTLT towards the eyes in Block 1 (*M* = .33, *SD* = .23) and Block 2 (*M* = .35, *SD* = .20) had a Bayes Factor of BF = .41 ($t(24) = - 1.23$, $p = .23$).

Paired Bayesian *t*-tests were also calculated to further assess the Pronunciation x AOI interaction by comparing the mean PTLT scores for each pronunciation type. correct versus consonant PTLT scores had a Bayes Factor of BF = 36.72 ($t(24) = -3.77$, $p < .001$), and correct versus vowel PTLT scores had a BF = 6.28 ($t(24) = -2.94$, $p = .01$), which both indicate support for the alternative hypothesis. In contrast, consonant versus vowel PTLT scores had a BF = .24 ($t(24) = .57$, $p = .58$), demonstrating support for the null hypothesis.

Vocabulary measures and face looking behaviour

To correct for multiple comparisons, a Bonferroni correction was applied to allow for 16 comparisons (.05/16 = .003). Infants had an average CDI comprehension of 11.42 words (*SD* = 11.42) out of 100, with all parents reporting that their infant knew the word *mummy*. No correlations were found between infants' listening preferences, (correct versus consonant, correct versus vowel, consonant versus vowel) and their CDI score ($p > .003$ in all cases, BF = .25 to .64). Infants were familiar with an average 11.08 (*SD* = 7.95) consonant tiers and 9.46 (*SD* = 5.49) vowel tiers. The proportion of unique consonant tiers (*M* = .98, *SD* = .03) was found to be significantly different to the proportion of vowel tiers (*M* = .91, *SD* = .12), $t(23) = 3.45$. $p = .002$. Again, no correlation was found between infants' listening preferences and either the proportion of consonant or vowel tiers they knew ($p > .003$ in all cases, BF = .25 to .47). Similarly, no correlation was found between infants' overall PTLT score (*M* = -.11, *SD* = .42) and either CDI and word list comprehension, or the number of consonant and vowel tiers they were familiar with ($p > .003$ in all cases, BF = .25 to .60). All correlations are listed in Appendix D.

The results of Experiment 5 differed from Experiment 4 in relation to both infants' overall looking time and their face scanning behaviour. Firstly, the total time infants spent focussing on the speaker varied based upon pronunciation type. The correct pronunciation and the vowel mispronunciation were both preferred over the consonant mispronunciation. However, no differences were found in attention to the speaker for the correct pronunciation and the vowel mispronunciation. This finding contrasts with Experiment 4, which demonstrated that infants did not differentiate between pronunciation types. Instead, this pattern of preferences is compatible with past research demonstrating a consonant bias in French-learning infants at 11 to 12 months (Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020). Additionally, such a familiarity preference differed from previous audiovisual research demonstrating that infants attended longer to audiovisually presented unfamiliar words (Weatherhead & White, 2017).

Infants' face scanning behaviour was found to be influenced by the pronunciation type. Overall, infants attended equally to the speaker's mouth and eyes, again replicating the typical looking patterns previously observed at this at this age (Lewkowicz & Hansen-Tift, 2012). Infants also looked more towards the speaker's eyes during the second half of the experiment. Furthermore, there was also a significant tendency for infants to attend less to the speaker's mouth when viewing both the consonant and vowel mispronunciations in comparison to the correct pronunciation. This suggests that infants treated both phonetic alterations comparably. Finally, CDI comprehension and the proportion of consonant and vowel tiers understood from these words, did not impact infants' listening preferences. Thus, the presence of a consonant bias did not appear to be influenced by infants' lexical understanding.

5.3 Interim Discussion

By the end of the first year of life infants not only match audio and visual speech cues (Kuhl & Meltzoff, 1982; MacKain et al., 1983; Patterson & Werker, 2003), but their speech perception also benefits from the additional phonetic details provided by their simultaneous presentation. Specifically, audiovisual speech can assist infants with discriminating between non-native phonemes (Ter Schure et al., 2016), learning unimodal distributed phonemic contrasts (Teinonen et al., 2008), and segmenting speech (Tan et al., 2023). The experiments presented in Chapter Four explored whether presenting British English-learning 12-month-olds with audiovisual speech cues assisted their capacity to detect phonetic changes to familiar word forms and led to evidence of a consonant bias in word form recognition. In addition, based upon past findings showing that infants adapt their typical looking patterns towards a talking face when viewing a speaker produce unfamiliar speech (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015), or when the audio and visual signals are mismatched (Mercure et al., 2019, 2022; Pejovic et al., 2020; Tomalski et al., 2013), it was examined whether infants' attention to a speaker's eyes and mouth differed between correct pronunciations and phonetic alterations.

In the first of two eye-tracking studies, Experiment 4 found that 12-month-old British English-learning infants looked equally towards videos of a speaker when they articulated lists of familiar word forms, irrespective of whether the words were pronounced correctly, with consonant alterations, or vowel alterations. Such preferences were again found to be irrespective of lexical factors. This finding replicated the results of Experiment 2b, in which infants did not differentiate between the same phonetic changes when they were presented auditorily. Furthermore, no significant differences were found in infants' attention to the speaker's eyes and mouth when viewing each pronunciation type. Instead, infants looked equally to the speaker's eyes and mouth during each list, a looking pattern that has previously been found in infants when viewing fluent speech at this age (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015). This further indicated that infants treated both consonant and vowel phonetic alterations of words comparably to their correct pronunciation. As with previous results in the present thesis (Experiments 2 and 3), these findings deviate from the consonant bias found in word form recognition (Poltrock & Nazzi, 2015) and own name recognition (Von Holzen & Nazzi, 2020) tasks in French-learning infants at this age. Thus, Experiment 4 appeared to support the finding that British English-learning infants treat consonant and vowel changes to lists of familiar word forms equally at 12 months even when they are presented audiovisually.

However, the results of Experiment 5 complicate this conclusion. When presented with a speaker pronouncing phonetic variations of "mummy", infants did not differ in their overall looking time to the face for a correct pronunciation and a vowel mispronunciation. In contrast, infants did look significantly longer when the face articulated both the correct pronunciation and the vowel mispronunciation compared to the consonant mispronunciation. This suggests that infants displayed a preference for vocalic over consonant information in recognising a single familiar word that was presented audiovisually. Thus, this finding is consistent with previous research demonstrating a consonant bias in French-learning infants at the end of their first year (Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020). Finally, infants' preferences were not correlated with lexical factors.

The pronunciation type also appeared to modulate infants' selective attention to areas on the speaker's face in Experiment 5. Infants paid equal attention to the eye and mouth regions of the face, although they looked significantly less at the mouth during the second half of the experiment in comparison to the first half. Importantly, infants looked significantly longer at the eyes for both phonetic mispronunciations in comparison to correct pronunciations. Thus, infants were relying less on the information provided by the speaker's mouth for both consonant and vowel alterations. Furthermore, this result indicates that, even though infants were showing a preference for the consonant change over the vowel change, they appeared to detecting both of the phonetic changes presented to them.

The potential explanations for these findings will be addressed in the General Conclusion. This final section will provide a summary of the results of both the head-turn preference and the eye-tracking experiments presented in the thesis, before considering the theoretical implications of these findings. It will then detail the limitations of these studies and discuss potential ideas for future directions to further the understanding of the acquisition of the consonant bias in British English-learning infants.

6 General Conclusion

6.1 Overview of key findings

Extensive research in adults has demonstrated that a consonant bias is found in most languages during lexical processing tasks (Nazzi & Cutler, 2019). However, such a preferential role of consonants over vowels does not appear to be an innate bias since cross-linguistic differences are found in its developmental course (Nazzi et al., 2016). For example, whilst data on French-learning infants unequivocally show a consonant bias emerging between 8 and 11 months (Nishibayashi & Nazzi, 2016; Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020), evidence regarding English-learning infants is less straightforward, with studies suggesting a possible consonant bias at 15 months (Mani & Plunkett, 2007) and 30 months (Nazzi et al., 2009), but no preference for either consonants or vowels at 12, 16/18, and 23/24 months (Floccia et al., 2014; Mani & Plunkett, 2007, 2010). Although the lexical regularities (Keidel et al., 2007) and the acoustic-phonetic information (Floccia et al., 2014) found in an infant's native language have been hypothesised to impact the age in which a consonant bias is observed, the existing evidence does not fully allow for a conclusion of the relative contribution of these factors. Thus, additional cross-linguistic analysis is necessary to further discern the role that both lexical and acoustic-phonetic factors may have in acquiring a consonant bias.

The primary aim of this thesis was to extend previous research into the development of the consonant bias by examining the phonetic specificity of British English-learning infants' word form recognition during their first year of life. Chapter Two detailed findings from a series of head-turn preference studies investigating whether consonant and vowel mispronunciations differentially impacted auditory-only word form recognition at 5, (Experiment 1), 11, (Experiment 2), and 12 (Experiment 3) months. Chapter Four then described the results from two eye-tracking studies in 12-month-olds, exploring whether a consonant bias was found when infants were presented with audiovisual stimuli and if consonant and vowel mispronunciations modified their face scanning behaviour (Experiment 4 and 5). Additionally, the impact of the acoustic factor of energy on word form recognition (Experiment 1) and whether lexical factors predicted the presence of a consonant bias (Experiments 2 to 5) were measured.

6.1.1 Head-turn preference research (*Experiments 1 to 3)*

Experiment 1 examined whether phonetic changes, as well as the acoustic factor of intensity, impacted word form recognition in British English-learning 5-month-olds. At this age, infants are typically beginning to recognise, and even comprehend, some frequently heard words (e.g., Bergelson & Swingley, 2012; Mandel et al., 1995; Tincoff & Jusczyk, 1999). They are also becoming attuned to their native vowels (e.g., Polka & Werker, 1994), whilst their perception has not yet narrowed to their native consonants (e.g., Werker & Tees, 1984). Previous research using the head-turn preference procedure by Delle Luche et al. (2017) had suggested that own name recognition was not phonetically specified in British English-learning infants at 5 months, such that a one-feature consonant or vowel mispronunciation was not differentiated from a correct pronunciation. However, post-hoc analysis revealed that a correct pronunciation was attended to longer when it began with a short amplitude-rise-time consonant than when it began with a longer amplitude-rise-time consonant. This indicated that British English-learning 5-month-olds' word recognition was potentially impacted by energy cues. Such findings contrasted with evidence from French-learning infants who displayed a vowel bias when presented with phonetic mispronunciations of their name at 5 months (Bouchon et al., 2015; Von Holzen & Nazzi, 2020), and a potential increased preference for large acoustic differences between vowels (Bouchon et al., 2015; but see Von Holzen & Nazzi, 2020).

Using the head-turn preference procedure, infants in Experiment 1 were presented with the correct pronunciation of the word *mummy* (/mʌmi/), a word they were likely to recognise at this age (Tincoff & Jusczyk, 1999), and either a phonetic or intensity alternation. Providing infants with variations of the same word, rather than their own name as in previous research (Bouchon et al., 2015; Delle Luche et al., 2017; Von Holzen & Nazzi, 2020), allowed for a controlled manipulation of both the phonetic and intensity alternations of the stimuli. Phonetic changes consisted of either an onset continuant-continuant consonant change (*nummy*; /nʌmɪ/), an onset continuant-stop consonant change (*bummy*; /bʌmi/), or an onset vowel change (*memmy*; /mɛmi/). Intensity changes were comprised of a 5dB energy increase to either the word's onset consonant (***m**ummy*; /**m**ʌmi/) or the word's onset vowel (*m**u**mmy*; /m**ʌ**mi/). Increased orientation times toward one pronunciation type over the other was interpreted as infants discriminating between auditory stimuli.

In relation to phonetic changes, the results of Experiment 1 suggested that British English-learning 5-month-olds' word form recognition is equally impacted by consonants and vowels. No difference was found in listening times to the correct pronunciation of the word *mummy* or the phonetic change it was presented with. This result supported previous research demonstrating that British English-learning 5-month-olds' own name recognition was not phonetically specified at this age (Delle Luche et al., 2017).

Furthermore, it contrasted with the vowel bias shown at this age in French-learning infants' own name recognition (Bouchon et al., 2015; Von Holzen & Nazzi, 2020). Thus, British English-learning infants not only appeared to fail to detect phonetic changes in familiar word forms but also differed from the vowel bias shown in their French-learning peers.

Similarly, Experiment 1 found that infants' word form recognition was not impacted by the controlled change of intensity. Listening times were equal to the unaltered pronunciation of *mummy* and both the increased energy made to the word's initial consonant and vowel. This result contrasted with the post-hoc findings of Delle Luche et al. (2017), in which own name recognition appeared to be driven by the energy information found in consonants. When specifically controlling for this energy change in the presentation of a single familiar word form, infants here did not display a preference for this acoustic alteration. This provides evidence that British English-learning infants do not appear sensitive to energy differences on both consonants and vowels in their early word form recognition.

One further factor to consider is whether the position of the vowel change within the word *mummy* impacted the results found in Experiment 1. Past research in 5-month-old French-learning (Bouchon et al., 2015) and British English-learning (Delle Luche et al., 2017) infants presented own-name alterations that occurred on the onset consonant (e.g., Ben to Wen) or onset vowel (e.g., Alfie to Elfie). In Experiment 1, whilst the consonant change remained on the onset of the word presented to infants, the vowel change occurred on the second sound. However, there is evidence to suggest that this methodological difference did not impact the results in this experiment. Firstly, the absence of a vowel bias found in Experiment 1 was consistent with the findings reported in Delle Luche et al. (2017). Secondly, the stress on the word *mummy*, as well as the vowel change to memmy, occurs on its first syllable. Past research has shown that such stress-initial words are likely to be highly salient to English-learning infants (Jusczyk et al., 1993). Thirdly, as demonstrated in Nishibayashi and Nazzi (2016), 6-month-old French-learning infants are also sensitive to vowel changes in the second sound of previously heard monosyllabic words (e.g., /ti/ versus /te/) but are not sensitive to onset consonant changes (e.g., /ny/ versus /by/). Together, this indicates that the position of the consonant and vowel changes within the familiar word were unlikely to have negatively impacted infants' abilities to detect the phonetic alterations presented to them.

Experiment 2 consisted of a series of three head-turn preference studies examining the consonant bias in British English-learning infants' word form recognition at the slightly older age of 11 months. At this age, infants' perception has now typically become attuned to their native vowels and is narrowing to their native consonants (e.g., Polka & Werker, 1994; Werker & Tees, 1984). They also show a preference for lists of familiar word forms over lists of unfamiliar words (Hallé & de Boysson-Bardies, 1994; Swingley, 2005; Vihman et al., 2004). French-learning infants already appear to display a consonant bias by this age (Nishibayashi & Nazzi, 2016; Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020). In particular, Poltrock and Nazzi (2015), having confirmed a preference for familiar over novel word forms, found that French-learning 11-month-olds preferred to listen to vowel alterations over onset consonant alterations of the same familiar word forms. Due to infants' preferences for familiar words at this age, this finding suggested a reliance on consonant information over vowel information in word form recognition at 11 months. Experiment 2 aimed to replicate this result in British English-learning 11-month-olds. Additionally, the relationship between lexical factors and infants' preferences was measured. This was examined through parents completing the 100-word Oxford Short Form CDI (Floccia et al., 2018; Hamilton et al., 2000) to determine the number of words infants understood, as well as calculating the number of unique consonant and vowel tiers that infants knew from these words.

Experiment 2a first examined British English-learning 11-month-olds' listening times to lists of correctly pronounced words that were likely to be recognised by them (e.g., *bottle*, *cuddle*, *mummy*) compared to lists of pseudowords (e.g., *puckle*, *gannel*, *nebby*). Past research suggested that infants would display a familiarity effect, such that words would be attended to longer than nonwords (Hallé & de Boysson-Bardies, 1994; Swingley, 2005; Vihman et al., 2004). The results confirmed this prediction, with infants showing a preference for words over pseudowords. This effect was found to be irrespective of parental reports of the number of words their infants knew, and to the phonological properties of infants' individual vocabularies in terms of consonant and vowel tiers. Thus, Experiment 2a added additional support to the literature indicating that, by 12 months, infants are beginning to recognise some frequently heard word forms (Hallé & de Boysson-Bardies, 1994; Swingley, 2005; Vihman et al., 2004).

Given the result of Experiment 2a, following the logic in Poltrock and Nazzi (2015), any preference for lists of either vowel or consonant mispronounced words would indicate that infants consider one phonetic variation more similar to correctly pronounced familiar words. This was examined in Experiment 2b, where a new group of infants were presented with onset consonant mispronunciations (e.g., *pottle*, *guddle*, *nummy*) versus vowel mispronunciations (e.g., *buttle*, *caddle*, *memmy*) of the familiar word forms presented in Experiment 2a. If infants behaved in the same way as their French-learning counterparts in Poltrock and Nazzi (2015), then a preference for lists of word forms with a vowel mispronunciation should emerge, establishing a consonant bias. However, British English-learning infants did not show such a greater dependence on consonants over vowels when recognising familiar word forms. Instead, no significant differences were found in listening time to either list. Furthermore, both the number of words infants were reported to understand, and the number of phonological tiers contained in these words, did not predict the presence of a consonant bias. This suggested that, irrespective of lexical factors, the infants were treating both consonant and vowel mispronunciations as being equally important, or equally unimportant.

This conclusion was further supported by the findings of Experiment 2c, which was run alongside Experiment 2b. In this study, infants were presented with medial consonant alterations (e.g., *boddle*, *cunnel*, *mubby*) and vowel mispronunciations (e.g., *buttle*, *caddle*, *memmy*) of the familiar words used in Experiment 2a. Although the consonant change was in the less salient stressed syllable of the word, and as such would be less recognisable to infants, the results nevertheless confirmed that infants did not differentiate between the phonetic alterations made to the words. Additionally, lexical measures were again not found to be linked to whether infants showed a consonant bias in this task.

Overall, Experiment 2 suggested that British English-learning 11-month-old infants did not demonstrate a consonant bias in a head-turn preference word recognition task. However, the lack of a bias for one type of mispronunciation over another could not be taken as a firm indication that infants did process each type of change equally. Indeed, the Bayesian analysis in Experiments 2b and 2c indicating inconclusive evidence of this null hypothesis. Experiment 3 aimed to address this by using a between-participant design to evaluate directly whether a change from a correctly pronounced familiar word (*mummy)* to a consonant alteration (*nummy*) would produce a similar preference compared to a change from a correctly pronounced familiar word to a vowel-modified version (*memmy*). To improve the likelihood of observing a consonant bias, older 12-month-olds were tested. This allowed for infants to have an increased consolidation of their native consonants (Polka & Werker, 1994), as well as a potentially slightly larger receptive vocabulary. Presenting infants with only one word and a mispronunciation also reduced infants' working memory load by reducing the number of items they heard.

The results of Experiment 3 appeared to be clear cut. Infants showed an overall preference for the correctly produced familiar word, which was similar in both consonant alteration and vowel alteration conditions. Thus, as in Experiment 1 and 2, there was no evidence of a consonant bias in British English-learning infants' word recognition during their first year of life. Infants' preferences again did not correlate with the size of their lexicons or the knowledge of unique consonant or vowel tiers in their vocabulary. This further indicated that regularities in the early lexicon did not seem to drive the emergence of a processing bias.

Taken together, the findings from Experiments 1 to 3 suggested that, by the end of their first year of life, a consonant bias is not found during lexical processing in British English-learning infants. Instead, infants learning this language display an equal sensitivity to consonants and vowels. The lack of a consonant bias in British English-learning infants in a head-turn preference word form recognition task contrasts with similar research conducted in infants learning other languages. Studies on French-learning infants have provided robust evidence for the preferential role of consonants in lexical processing tasks from an early age. Developing from an initial vowel bias present until around 6 months (Bouchon et al., 2015; Nishibayashi & Nazzi, 2016), French-learning infants have been found to demonstrate a consonant bias from 11 months in familiar word recognition tasks (Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020), and even from 8 months in word segmentation tasks (Nishibayashi & Nazzi, 2016). This early consonant bias maintains in French-learning infants, so that, by the time they are between 16 and 20 months, consonants have a privileged role over vowels in word-learning tasks (Havy & Nazzi, 2009; Nazzi, 2005; Nazzi & Bertoncini, 2009; Nazzi & New, 2007). Similarly, Italian-learning infants appear to follow a comparable development trajectory to French-learning infants, showing a higher sensitivity to vowels over consonants at 6 months, before showing a reverse pattern at 12 months (Hochmann et al., 2011; Hochmann et al., 2017).

Instead, Experiments 1 to 3 appear to add further support to previous research suggesting that infants learning British English (Delle Luche et al., 2017; Floccia et al., 2014), Danish (Højen & Nazzi, 2016), and tone languages (Chen et al., 2021; Singh et al., 2015; Wewalaarachchi & Singh, 2020) show cross-linguistic differences to French- and Italian-learning infants in the acquisition of a consonant bias. In relation to British English, infants learning this language can detect their own name versus a phonetically dissimilar name at 5 months but are unable to do so when versus either a consonant or vowel mispronunciation of their name (Delle Luche et al., 2017). Older infants also fail to show a differential processing of consonants and vowels, displaying an equal sensitivity to both phonetic types. Specifically, 12-, 18-, and 24-month-old British English infants are equally impacted by both consonant and vowel mispronunciations of familiar words, with only 15-month-old infants showing a sensitivity to consonant over vowel changes (Mani & Plunkett, 2007, 2010). This finding extends to word- learning tasks, with the acquisition of word pairings that vary by either one consonant or one vowel comparably learned in both 16- and 24-month-old infants (Floccia et al., 2014), with a consonant bias not appearing in this task until 30 months (Nazzi et al., 2009). Experiments 1 to 3 add further weight to these findings, showing that 5-, 11-, and 12-month-old British English-learning infants are equally sensitive to consonant and vowel changes in word form recognition tasks.

6.1.2 Eye-tracking research (*Experiments 4 and 5)*

Building on the head-turn preference findings, Experiment 4 assessed whether providing British English-learning 12-month-olds with visual information enhanced their ability to differentiate between phonetic alterations of familiar word forms. There were several reasons to predict that viewing a speaking face would lead to a consonant bias being found. Firstly, infants are sensitive to the temporal correspondence between audio and visual speech cues from very early on in development, demonstrating intersensory matching of auditory phonemes (Aldridge et al., 1999; Kuhl & Meltzoff, 1982; Patterson & Werker, 2003) and fluent speech (Dodd, 1979; Kubicek et al., 2014) to their correct visual articulation. As with auditory-only phonemes, this ability narrows to an infants' native audiovisual phonemes and language during their first year of life (Kubicek et al., 2014; Pons et al., 2009). Secondly, infants can successfully integrate audio and visual speech signals, as demonstrated by the impact of the McGurk effect on their speech perception (Bristow et al., 2008; Burnham & Dodd, 2004; Rosenblum et al., 1997). Thirdly, audiovisual speech can boost infants' speech perception, such as assisting their capacity to discriminate between non-native phonemes following phonological acquisition (Ter Schure et al., 2016) and acquire unimodal distributed phonemic contrasts (Teinonen et al., 2008). Finally, 12- to 13-month-olds can differentiate between audiovisual familiar and unfamiliar words, displaying a novelty preference when presented with such stimuli (Weatherhead & White, 2017).

Experiment 4 also investigated whether infants' face scanning behaviour differed based upon the pronunciation type presented to them. While adults will typically focus on a speaker's eyes, infants' attention towards a speaker's face is subject to two developmental shifts between 4 and 12 months. Beginning with a focus on the eyes of a speaker at 4 months, this shifts to the speaker's mouth by 8 months, before moving back to an adult-like attention towards the eyes, with an equal attention to a speaker's eyes and mouth at 12 months (e.g., Lewkowicz & Hansen-Tift, 2012). However, this typical looking behaviour is disrupted during the first year of life when infants are presented with an unfamiliar language or incongruent audio and visual syllables, with attention found to increase towards either the mouth or the eyes depending on the age of infants and the type of stimuli (Lewkowicz & Hansen-Tift, 2012; Mercure et al., 2019, 2022; Pejovic et al., 2020; Pons et al., 2015; Tomalski et al., 2013). Thus, there was reason to expect that, if infants detected the mispronunciations of a familiar word, then they would alter their looking pattern towards a speaker's face.

Accordingly, Experiment 4 presented British English-learning 12-month-olds with sequences of videos of a speaker pronouncing a series of familiar words, spoken correctly, with an onset consonant mispronunciation, and with a vowel mispronunciation. The words presented to infants were the same as those heard in Experiment 2a, whereas the phonetic alterations were identical to those heard in Experiment 2b. An eye-tracker was used to measure how long infants attended to the speaker during each list, as well as their attention towards the speaker's mouth and eyes. The findings again demonstrated that infants did not differentiate between correct pronunciations and phonetic alterations, with overall looking times to the face equal for each list. Similarly, the pronunciation type did not impact looking patterns towards areas on the speaker's face. Instead, infants attended equally to the eyes and mouth in all conditions. Additionally, linguistic measures did not correlate with either the infants' pronunciation preference or the amount of time they spent attending to the speaker's eyes or mouth. Therefore, it appeared that, even with the additional phonetic cues provided by viewing a speaker's mouth, infants still did not show a consonant bias in word form recognition.

In a final study, Experiment 5 simplified the task further for 12-month-olds, presenting infants with audiovisual stimuli of a single word (once more, *mummy*) along with an onset consonant (*nummy*) and vowel (*memmy*) mispronunciation. Overall looking times to the face again showed that infants attended equally to an audiovisually presented correct pronunciation and vowel mispronunciation. However, the amount of time spent looking at the speaker was found to be longer for both the correct pronunciation and the vowel mispronunciation compared to the consonant mispronunciation. This result was comparable to findings in French-learning 11-month-olds who display a consonant bias when presented with phonetic alterations made to lists of familiar words (Poltrock & Nazzi, 2015) or their own name (Von Holzen & Nazzi, 2020) in head-turn preference tasks. Thus, it appeared that British English-learning infants were displaying a consonant bias when presented with audiovisual phonetic alterations of a single word.

Infants in Experiment 5 also looked equally towards the eyes and mouth of the speaker overall. However, they were found to spend significantly longer looking at the speaker's eyes when viewing both the consonant mispronunciation and the vowel mispronunciation. This finding indicated that infants detected both phonetic changes made to the familiar word. Furthermore, it suggested that infants relied less on the information provided by the speaker's mouth when viewing such mispronounced articulations. Finally, lexical measures once again did not correlate with infants' preferences of face scanning behaviour. Thus, the number of words and phonological tiers infants were reported to know did not impact the presence of a consonant bias.

Experiment 5 suggested that audiovisual stimuli appeared to facilitate a consonant bias in British English-learning 12-month-olds when viewing phonetic mispronunciations of the word *mummy*. This result contrasted with the previous research presented in this thesis. Firstly, it differed from Experiment 3, where infants preferred listening to a correctly pronounced version of *mummy* over both a consonant change to *nummy* and vowel change to *memmy* in the absence of visual cues. One explanation for divergent findings is that audiovisual speech provides phonetic and temporal information that is not supplied by the auditory signal alone (Lalonde & Holt, 2015). The extra phonetic details provided by the visual articulations of the phonetic changes may have assisted infants' abilities to detect such mispronunciations in Experiment 5, when they were unable to do so in a head-turn preference task in Experiment 3. This is supported by evidence that providing concurrent audio and visual signals can improve infants' speech perception during their first year of life compared to when they are presented with an auditory signal alone (e.g., Teinonen et al., 2008; Ter Schure et al., 2016).

An alternative explanation for the findings in Experiment 5 is that the temporal information provided by the audiovisual stimuli may have improved infants' ability to detect the mispronunciations. Mouth movements occur between approximately 60 and 300ms before the audio signal (Chandrasekaran et al., 2009; Schwartz & Savariaux, 2014). The amplitude envelope of the auditory signal is also louder as the mouth opens and quieter when the mouth closes (Chandrasekaran et al., 2009). Such cues can improve temporal expectancy by reducing the uncertainty of when the audio signal will begin (Lalonde & Werner, 2021). Previous research suggests that temporal information is beneficial to infants in speech perception tasks. For instance, providing cues relating to the onset of speech can enhance infants' abilities to discriminate speech in noise (Lalonde & Werner, 2019). Similarly, viewing an oscilloscope pattern depicting the temporal envelope of an audio signal can assist infants word segmentation (Hollich et al., 2005). In relation to the present results, the ability for infants to detect the phonetic alterations in Experiment 5 may have been assisted by such visual cues marking when they should attend to the speech signal. Attending to the start of each token may have been particularly important as the consonant and vowel changes occurred in the first syllable of each mispronunciation. Thus, infants in Experiment 5 may have particularly benefited from the temporal information provided by a speaking face when presented with phonetic alterations.

However, the present results do not allow for the conclusion that the addition of temporal information alongside the audio signal led to a consonant bias being displayed, given that audiovisual speech provides details concerning both temporal and phonetic cues. Future research should therefore examine the role that temporal cues may have in assisting British English-learning infants in their ability to detect mispronunciations. For example, the impact of temporal cues could be explored by providing infants with the audio signal and visual cues that contain only temporal information. For instance, infants could be shown an image of the speaker with their mouth open at the onset of the word's auditory signal, and an image of the speaker with their mouth closed at the offset of the word's auditory signal. If a consonant bias was found in British English-learning infants when only the temporal properties of the speech signal are boosted, it would suggest that additional temporal cues are more beneficial than phonetic cues in leading infants to show a preference for consonant over vowel mispronunciations.

The results of Experiment 5 also differed from Experiment 4, where infants did not show a consonant bias or change their looking patterns towards the speaker's eyes and mouth when presented with audiovisual lists of familiar word forms and their phonetic mispronunciations. There are several possible explanations for these conflicting findings. For example, presenting infants with audiovisual alternations of the word *mummy* in Experiment 5 reduced the working memory required for the task compared to the lists presented in Experiment 4. Such differences in the complexity of the task may explain why infants showed a familiarity effect for audiovisual stimuli in Experiment 5, compared to previous research demonstrating a novelty effect when viewing audiovisual familiar and unfamiliar word forms (Weatherhead & White, 2017).

Differences in the feature changes made to the onset consonants of the familiar words presented in Experiments 4 and 5 may have led to the dissimilar results. For consonants, the visual signal alone provides information concerning the place of articulation (e.g., /m/ to /n/). Conversely, cues relating to voicing (e.g., /d/ to /t/) are provided by the audio signal, with the manner of articulation (e.g., /b/ to /m/) conveyed by both the audio and visual signal (Ladefoged & Disner, 2012). Previous research has shown that 12- to 13-month-olds fail to detect voiced changes made to the onset consonants of audiovisually presented familiar words but do recognise more visually salient place changes (Weatherhead & White, 2017). Experiment 4 contained several types of phonological feature changes to consonants, which varied in their visual distinctiveness (4 place, 3 voice, 3 manner). As suggested in Weatherhead and White (2017), voiced changes may activate comparable phonemes (e.g., /t/ activating /d/) even though the audio signal does not match. Thus, viewing the initial phoneme /t/ before *-addy* may have led infants to recognise the word *daddy*. Similarly, viewing an onset voiced change may activate the lexical neighbourhood of visually consistent articulations, which could potentially assist infants with identifying a word based upon the remaining input. For instance, observing the visual onset phoneme /t/ may activate lexical candidates which share visually consistent onsets (e.g., *daddy*, *dummy, teddy*, *tickle*, *tummy*), leading to the recognition of the word *daddy* when audiovisually presented with *taddy*. Together, this suggests that infants' attention towards the consonant mispronunciations in Experiment 4 could have been driven by the voiced feature changes contained within the lists presented to them.

In contrast, the sole feature change presented in the consonant mispronunciation condition in Experiment 5 was made to the place of articulation at the onset of the word (i.e., *mummy* to *nummy*). Although this feature change was also presented in Experiment 4, it was likely to be particularly visually salient to infants when presented in isolation due to the visual signal providing information on the place of articulation. Consequently, this may have assisted infants with their detection of the mispronunciation. However, it is important to note that the visual onset of this mispronunciation may have potentially impacted infants' word recognition. Specifically, observing the onset consonant /n/ could lead to the activation of comparable visual phonemes (e.g., /d/, /t/), and subsequently activate visually consistent familiar word forms when followed by *-ummy* (i.e., *dummy, tummy*). Therefore, this may have hypothetically led some infants to treat this mispronunciation as the correct pronunciation of a familiar word form and increased their overall attention towards the speaker.

It is also necessary to consider that vowels are less visually recognisable than consonants, due to differing in their positioning of the tongue rather than their manner of articulation (Stevens & House, 1955). Nevertheless, the visual signal provides the most information concerning the roundness of vowels, followed by their height, and then finally their backness (Benguerel & Pichora-Fuller, 1982). Thus, the phonological feature changes made to the vowels in Experiment 4 (3 height, 7 backness) and Experiment 5 (1 backness) may have impacted how visually distinctive the vowel mispronunciations were. In particular, the vowel backness feature change from *mummy* to *memmy* in Experiment 5 may not have been as visually salient as the consonantal place change to *nummy*. This potentially could have influenced infants' abilities to detect the difference between the vowel change and the correct pronunciation. However, given that infants adjusted their face scanning behaviour in Experiment 5 towards both phonetic changes, they appeared to be sensitive to such alterations, irrespective of their visual distinctiveness.

However, further research is required to establish the role that the visual distinctiveness of consonants and vowels may have in infants displaying a consonant bias in Experiment 5. In particular, given that the audio signal did not lead to a consonant bias in British English-learning 12-month-olds (Experiment 3), it should be examined whether presenting the visual signal in isolation would still lead to such infants displaying a consonant bias. Specifically, infants could be presented with videos of a speaker silently articulating the correct and phonetic alterations of a familiar word. If infants still showed a preference for a correct pronunciation and a consonant mispronunciation over a vowel mispronunciation in the absence of any auditory information, this would indicate that the visual cues alone were leading to a consonant bias. Importantly, such a finding would demonstrate that the consonant bias found in Experiment 5 was due to the visual properties of the phonetic changes presented to infants. Conversely, if infants did not show a consonant bias in the absence of the auditory signal, this would suggest that it was the combination of auditory and visual signals that led to the consonant bias being shown by infants in Experiment 5.

In relation to infants' attention to the eyes and mouth of the speaker, both Experiments 4 and 5 found that infants looked equally towards the two areas of interest when viewing each pronunciation type. This provided further evidence that, when viewing a native talker, 12-month-olds divide their looking time between the eyes and mouth (Lewkowicz & Hansen-Tift, 2012). However, when presented with a single repeating word form in Experiment 5, infants adjusted their looking behaviour towards both phonetic changes compared to the correct pronunciation. This finding adds to recent research demonstrating that the typical looking patterns towards a speaker's face during the first year of life are disrupted when attending to either a novel non-native language (Lewkowicz & Hansen-Tift, 2012; Pons et al., 2015) or incongruent audio and visual signals (Mercure et al., 2019, 2022; Pejovic et al., 2020; Tomalski et al., 2013). Additionally, this result indicates that infants were still appearing to detect both types of phonetic change. Thus, whilst infants displayed a consonant bias in their preference for mispronunciations in an audiovisual setting, they were still demonstrating an ability to recognise both consonant and vowel changes.

Furthermore, the change in looking patterns to mispronunciations in Experiment 5 was found to be an increase in attention towards the speaker's eyes and a reduction in attention to their mouth. Mercure et al. (2022) found a similar change in attention to the face in 7- to 10-month-olds when they viewed incongruent audio and visual signals. They proposed that such a shift in focus to the speaker's eyes may reflect a mature face scanning pattern in which infants pay less attention to unhelpful mouth movements. This explanation may be congruent with the present findings. Infants may have detected that the information provided by the speaker's mouth was unhelpful in recognising a familiar word, and thus focussed more on the speaker's eyes. Indeed, there is evidence to suggest that 6- to 9-month-olds who shift their attention away from unhelpful visual cues during incongruent audiovisual trials exhibit better language abilities between 14 and 16 months compared to infants who do not (Kushnerenko et al., 2013). However, research evaluating where on a speaker's face infants attend when hearing a familiar word and a novel word will be required to further evaluate this conclusion.

6.2 Theoretical implications

The data presented in this thesis allow further inferences to be made regarding the origin of the functional asymmetry between consonants and vowels. Recall that three developmental hypotheses concerning the acquisition of the consonant bias have been proposed. Of these three separate scenarios, one predicts that infants will show a consonant bias from birth (initial bias hypothesis), whereas two predict cross-linguistic differences (lexical hypothesis and acoustic-phonetic hypothesis).

The initial bias hypothesis (Nespor et al., 2003) suggests that the processing of consonants and vowels as separate linguistic categories is innate. An initial consonant bias would help infants to acquire the lexicon of their native language, whereas an initial vowel bias would assist infants with learning grammatical and prosodic information. Furthermore, neither developmental nor cross-linguistic variations should be present in infants acquiring the consonant bias. However, past findings in British English- (Delle Luche et al., 2017; Floccia et al., 2014), French- (Bouchon et al., 2015; Nishibayashi & Nazzi, 2016; Von Holzen & Nazzi, 2020), Italian- (Hochmann et al., 2011, 2017) and Danish-learning (Højen & Nazzi, 2016) infants strongly suggest that the consonant bias is not innate. The present findings add further support to this, with British English-learning infants not displaying a consonant bias at 5, 11, and 12 months in an auditory word recognition task and only appearing to show a preference for vowel over consonant information at 12 months when provided with additional phonetic cues. Thus, the results here substantiate the idea that the role of consonants in lexical processing emerges during development due to language experience (Nazzi et al., 2016), which favours an explanation based on acquisition of the native lexicon (Keidel et al., 2007) and/or to the acoustic–phonetic properties of the native language (Floccia et al., 2014).

The lexical hypothesis (Keidel et al., 2007) is based upon the finding that a word (e.g., casino) is around 40% more likely to be recognised from its consonant tier information (e.g., c-s-n) compared to its vowel tier information (e.g., a-i-o). This hypothesis suggests that the consonant bias is acquired once infants have learned that consonants are statistically more informative than vowels in word recognition tasks. This would occur at some point during an infant's second year, when their lexicon and/or their understanding of the distributional structure of the consonants and vowels in their native language reaches an undetermined amount. Furthermore, how informative the consonants and vowels of a given language are could lead to cross-linguistic differences in the age in which a consonant bias emerges. If this hypothesis is correct, one would expect that the size of an infant's growing lexicon, and/or the proportion of unique consonant tiers versus vowel tiers they know, would predict the emergence of the consonant bias. This was indexed in the present thesis by a preference for vowel-changed words over consonant-changed words (in Experiments 2b and 2c), or by a preference for correctly pronounced words over consonant or vowel mispronunciations (Experiments 3, 4, and 5). However, the results from such comparisons failed to confirm these predictions. In all cases, no correlation was found between the number of words infants were reported to understand and their listening preferences. This suggested that the size of their lexicon did not impact the presence of the consonant bias at 11 or 12 months. Similarly, although each experiment found that infants knew significantly more unique consonant tiers than unique vowel tiers, the proportion of both unique consonant and vowel tiers was not correlated with their listening preferences. This finding indicates that the structure of the infants' lexicon did not impact whether they displayed a consonant bias in these experiments. This adds further support to previous evidence suggesting that lexical factors do not modulate the presence of a consonant bias in French-learning 11-month-olds (Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020). Thus, a purely lexical hypothesis does not appear to account for the cross-linguistic differences found in the development of the consonant bias.

The acoustic-phonetic hypothesis (Floccia et al., 2014) also suggests that the consonant bias is learned, but through the exposure to the acoustic-phonetic properties of the consonants and vowels of a language, instead of through exposure to its lexical regularities. According to this hypothesis, the increased salience, periodicity and stability of vowels over consonants provides an initial advantage to an infant in processing vowels, with the consonant bias developing before their first birthday. This switch between an initial vowel bias and a subsequent consonant bias may occur as consonants are processed more categorically than vowels, which could highlight to infants that they are more dependable when both recognising and learning new words (Hochmann et al., 2011). It may arise due to consonant categories appearing to be more discriminable than vowel categories (Bouchon et al., 2015) or because of the emerging capacity to process fine temporal information (Werner et al., 1992). Such factors could help infants learn to pay more attention to consonants during lexical processing. Within this perspective, cross-linguistic differences due to differential consonantal and vocalic properties between languages may be found in the development of the consonant bias.

Indeed, the phonological properties of the French and English languages may help to explain the cross-linguistic differences in the reliance of consonants and vowels in past research in French-learning infants and the findings reported in the present thesis. French has a syllable-timed rhythm, phrase-final lengthening, and contains mainly steady-state vowels. Confronted with this, French-learning infants may initially focus on the vowels of their native language, but, with the expansion of the lexicon and the improvement of temporal resolution abilities, would develop a bias for giving more weight to differences between consonants. This would ultimately lead to the development of a consonant bias by the end of the first year of life (Bouchon et al., 2015; Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020). In comparison, British English-learning infants are exposed to a different range of acoustic information, with reduced or fully-realised vowels, variable lexical stress, ambisyllabicity and regular diphthongisation. These complex vowel properties may not drive infants' attention away from vowels to focus on consonants as early as found in French-learning infants. This may result in British English-learning infants facing a harder task to develop a functional distinction between consonants and vowels, as demonstrated in word form recognition here at 5, 11, and 12 months in Experiments 1 to 4 and in previous research before the age of 30 months across various lexical tasks (Delle Luche et al., 2017; Floccia et al., 2014; Mani & Plunkett, 2007, 2010). Therefore, even though infants can show a consonant bias in word recognition during infancy, namely at 15 months in an IPL task (Mani & Plunkett, 2007), and at 12 months when viewing the audiovisual mispronunciations of a single word (Experiment 5), such a lexical processing bias does not appear to be fully functional until later in development.

The developmental impact of a later acquisition of a fully operational consonant bias in British English-learning bias nevertheless remains unclear. British English-learning infants can clearly detect familiar word forms during their first year of life. For example, although they do not differentiate between correctly pronounced and one-feature phonetic changes of their own name at 5 months, they are still able to recognise their name in comparison to a phonetically dissimilar name containing the same number of syllables and stress pattern (Delle Luche et al., 2017). At 11 months, they also show a preference for frequently heard word forms over unfamiliar pseudowords (Experiment 2a), even if they do not yet show a consonant bias when listening to consonant and vowel alterations of the same familiar word forms (Experiments 2b and 2c). Similarly, at 12 months, they can match a familiar word to a visual referent, with their recognition equally impacted by consonants and vowel mispronunciations (Mani & Plunkett, 2010). However, there is evidence that British English-learning infants' lexical development, as measured by CDI, is slower during their first two years of life in comparison to infants learning other languages (e.g., Bleses et al., 2008). In fact, their vocabulary growth is very similar to that of Danish children, who are faced with the difficult task of learning a language with a highly complex vowel system. Given that the consonant bias is proposed to optimise lexical processing, the equal sensitivity to consonants and vowels found in British English-learning infants until 30 months may contribute to such delayed word recognition and learning abilities.

Finally, it should also be noted that phonological and lexical acquisition do not occur in isolation, with one likely to impact the development of the other (e.g., Yeung et al., 2014; Yeung & Nazzi, 2014; Yeung & Werker, 2009). As proposed by Nazzi et al. (2016), the lexical and acoustic-phonetic hypotheses may have a combined influence on the development of a consonant bias in lexical processing tasks. Consequently, infants may learn that the acoustic-phonetic properties of consonants and vowels differ in relation to cues for word recognition, as proposed by the acoustic-phonetic hypothesis. However, this will occur alongside infants obtaining the statistical knowledge of their native lexicon. Further research should attempt to investigate the combined impact of both lexical and acoustic-phonetic factors on the development of the consonant bias.

6.3 Methodological limitations and future directions for research

Experiments 1 to 3 used the head-turn preference procedure as a behavioural measure to examine infants' preferences for phonetic mispronunciations of familiar word forms. Although this paradigm has previously been successfully used to demonstrate the consonant bias in French-learning infants (e.g., Bouchon et al., 2015; Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020), a more sensitive measure may be required in British English-learning infants due to the harder task they may have with recognising phonetic alterations. Specifically, electrophysiological and neuroimaging techniques could be used to examine whether British English-learning infants display a consonant bias in word form recognition. For example, the ERP method has the advantage over the head-turn preference paradigm in that it does not require any behavioural change from the infant. Instead, infants passively attend to speech stimuli while their EEG is recorded. Although past research has used ERP to examine word recognition in infants (e.g., Duta et al., 2012; Kooijman et al., 2005; Mani et al., 2012; Thierry et al., 2003; Von Holzen et al., 2018), no study has used this methodology to directly compare neural response to consonant and vowel mispronunciations in British English-learning infants during their first year of life. Thus, using such a method to examine whether British English-learning infants' word form recognition is disrupted by consonant and vowel alterations may be advantageous to study the development of the consonant bias.

Experiments 4 and 5 examined how long infants looked towards videos of a speaker saying words correctly, with a consonant change, and with a vowel change, whilst also measuring whether their attention towards the speaker's mouth or eyes varied during each pronunciation type. Due to the different dimensions of the eyes and mouth, the size of the AOI around each target differed. Specifically, the percentage of the speaker's face covered by each AOI was larger around the speaker's eyes (18.6%) than the speaker's mouth (9.6%). One methodological issue with this experimental design is that larger AOIs may have received more fixations by chance if infants attended randomly to the speaker's face. It is important to note that the results in Experiments 4 and 5 did not indicate that infants' attention to the face was random, with infants overall looking equally to each AOI. Indeed, this overall pattern of attention towards a speaker's eyes and mouth replicated results found in 12-month-olds in previous research in which comparable AOI sizes had been used (e.g., Lewkowicz & Hansen-Tift, 2012). Nevertheless, future research may wish to address the issue of different AOI dimensions by applying AOIs that are more comparable in size. For example, two equal AOIs could be applied to the speaker's face, in which the area around the eyes and mouth are the same dimensions (e.g., see Berdasco-Muñoz et al., 2019) or in which the speaker's face is bisected into equally sized upper and lower halves (e.g., see DeBolt & Oakes, 2023).

The present thesis failed to find a link between lexical factors and the presence of a consonant bias in British English-learning infants, supporting similar results in French-learning infants (Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020). However, as suggested by Von Holzen and Nazzi (2020), assessing an infants' lexicon size, such as through use of CDI tools, only considers the words an infant knows, rather than the words they are exposed to in their environment. The present thesis relied on such estimates, calculating an infant's vocabulary size, along with the phonological tiers they knew, based upon the 100-word Oxford Short Form CDI (Floccia et al., 2018; Hamilton et al., 2000). Thus, the experiments here only examined the relationship between a limited number of familiar words, as well as their associated phonological tiers, and the consonant bias. Recent innovations in non-invasive long-format speech environment recordings have allowed researchers to successfully examine the words infants are exposed to over the course of an entire day (e.g., Casillas & Cristia, 2019; Gilkerson et al., 2017; VanDam et al., 2016). From these recordings, automatic transcriptions could be used to calculate the number of unique consonant and vowel tiers infants heard during the day. Consonant tier and vowel tier proportion scores, calculated by dividing the number of each unique consonant and vowel tiers heard by the total number unique words heard, could then be measured against infants' preference for consonant and vowel mispronunciations in a head-turn preference task. This would establish whether the lexical structure of words in an infant's input or the words an infant understands impacts the development of the consonant bias in infants.

Additional experiments may also explore the acquisition of the consonant bias in infants learning American English. Several developmental differences have already been reported between British English- and American English-learning infants. To illustrate, comparisons of CDI data have revealed that American English-learning infants are significantly ahead of British English-learning infants during their first few years of life in both their word recognition and word production abilities (Fenson et al., 1994; Hamilton et al., 2000). Similarly, whilst American English-learning infants can segment words from fluent speech from 7.5 months (Jusczyk & Aslin, 1995), British English-learning infants are unable to do so until 10.5 months, and only when presented with exaggerated infant directed speech (IDS, Floccia et al., 2016). American English-learning infants are exposed to different acoustic cues in comparison to British English-learning infants. For example, whilst both dialects share the same number of consonants, American English has fewer vowel sounds compared to British English (for a review, see Deterding, 2004). Other dissimilarities include variations in lexical stress between the two dialects (Berg, 1999) and differences in the pronunciation of /r/ (Ladefoged & Disner, 2012). Furthermore, American English IDS contains more prosodic modification than British English IDS (Fernald et al., 1989). Together, such factors may result in American English-learning infants having an easier task in learning the functional biases associated with consonants and vowels during their first year than their British English-learning peers. Therefore, further research should investigate whether these within-language differences lead to American English-learning infants displaying a consonant bias by applying the same experimental methods used on French- and British English-learning infants.

Finally, the results of Experiment 5 suggested that 12-month-old British English-learning infants displayed a consonant bias when audiovisually presented with phonetic alternations of a single word form. Further to this finding, it would be beneficial to examine whether the effects of audiovisual speech on the consonant bias extend to different ages. Research beyond the scope of the present thesis is currently underway to examine whether 5-month-old British English-learning infants differentiate between the same audiovisual stimuli that were presented to 12-month-olds in Experiment 5. Due to the addition of visual speech appearing to facilitate a consonant bias in 12-month-olds, providing younger infants with access to the speaker's mouth may improve their ability to detect phonetic changes. This may consequently lead to evidence of either a consonant bias, or even an initial vowel bias as displayed in French-learning infants at this age (Bouchon et al., 2015; Nishibayashi & Nazzi, 2016; Von Holzen & Nazzi, 2020). It would also be interesting to explore whether the concurrent presentation of audio and visual speech signals impacts older British English-learning infants' abilities to detect mispronunciations between a word form and its visual referent. Such research may adapt previous experiments by Mani and Plunkett (2007, 2010), using an inter-modal preferential looking task to present infants with familiar objects on a screen that are audiovisually, rather than auditorily, labelled correctly or with a phonetic mispronunciation. The addition of the visual speech signal may again lead to a consonant bias being shown in such a task.

6.4 Overall conclusion

In summary, the overarching aim of this thesis was to explore the developmental acquisition of the consonant bias in British English-learning infants during their first year of life. Taken together, the results demonstrated that infants' word form recognition appeared to be equally reliant upon consonants and vowels at 5, 11, and 12 months when presented with an auditory word form recognition task. Only when the task was simplified and additional phonetic information was provided, did infants appear to both detect the phonetic changes of a familiar word form and show a preference for a vowel mispronunciation over a consonant mispronunciation. Overall, the findings reported in this thesis add additional support to past research demonstrating cross-linguistic differences in the age a consonant bias appears in word form recognition, with the present results differing from a clear vowel bias at 5 months (Bouchon et al., 2015; Von Holzen & Nazzi, 2020), and a consonant bias at 11 months (Poltrock & Nazzi, 2015; Von Holzen & Nazzi, 2020) in French-learning infants. However, the underlying mechanisms of these cross-linguistic differences remain unclear, given that neither acoustic (at 5 months) nor lexical (at 11 and 12 months) factors appeared to impact word form recognition. Thus, further research is needed to explore the role such acoustic-phonetic and/or lexical factors have in shaping the way infants learn the beneficial role of consonants in lexical processing tasks.

7 References

Aldridge, M. A., Braga, E. S., Walton, G. E., & Bower, T. G. R. (1999). The intermodal representation of speech in newborns. *Developmental Science*, *2*, 42-46. https://doi.org/10.1111/1467-7687.00052

Altvater-Mackensen, N., & Grossmann, T. (2016). The role of left inferior frontal cortex during audiovisual speech perception in infants. *NeuroImage*, *133*, 14-20. https://doi.org/10.1016/j.dcn.2018.10.002

Altvater-Mackensen, N., Mani, N., & Grossmann, T. (2016). Audiovisual speech perception in infancy: The influence of vowel identity and infants' productive abilities on sensitivity to (mis)matches between auditory and visual speech cues. *Developmental Psychology*, *52*(2), 191–204. https://doi.org/10.1037/a0039964

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology*, *92*(2), 339-355. https://doi.org/10.1348/000712601162220

Aslin, R. N., Jusczyk, P. W., & Pisoni, D. B. (1998). Speech and auditory processing during infancy: constraints on and precursors to language. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: cognition, perception, and language*. New York: Wiley.

Auer, E. T. (2002). The influence of the lexicon on speech read word recognition: Contrasting segmental and lexical distinctiveness. *Psychonomic Bulletin & Review*, *9*(2), 341-347. https://doi.org/10.3758/BF03196291

Baart, M., Vroomen, J., Shaw, K., & Bortfeld, H. (2014). Degrading phonetic information affects matching of audiovisual speech in adults, but not in infants. *Cognition*, *130*(1), 31–43. https://doi.org/10.1016/j.cognition.2013.09.006

Bahrick, L. E., Todd, J. T., Castellanos, I., & Sorondo, B. M. (2016). Enhanced attention to speaking faces versus other event types emerges gradually across infancy. *Developmental Psychology*, *52*(11), 1705–1720. https://doi.org/10.1037/dev0000157

Barenholtz, E., Mavica, L., & Lewkowicz, D. J. (2016). Language familiarity modulates relative attention to the eyes and mouth of a talker. *Cognition*, *147*, 100-105. https://doi.org/10.1016/j.cognition.2015.11.013

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241-251. https://doi.org/10.1017/S0021963001006643

Barry-Anwar, R. A., Burris, J. L., Estes, K. G., & Rivera, S. M. (2017). Caregivers and strangers: The influence of familiarity on gaze following and learning. *Infant Behavior and Development*, *46*, 46-58. https://doi.org/10.1016/j.infbeh.2016.11.005

Bastianello, T., Keren-Portnoy, T., Majorano, M., & Vihman, M. (2022). Infant looking preferences towards dynamic faces: A systematic review. *Infant Behavior and Development*, *67*, 101709. https://doi.org/10.1016/j.infbeh.2022.101709

Beauchemin, M., Gonzalez-Frankenberger, B., Tremblay, J., Vannasing, P., Martínez-Montes, E., Belin, P., Béland, R., Francoeur, D., Carceller, A. M., Wallois, F., & Lassonde, M. (2011). Mother and stranger: an electrophysiological study of voice processing in newborns. *Cerebral Cortex*, *21*(8), 1705-1711. https://doi.org/10.1093/cercor/bhq242

Benavides-Varela, S., Hochmann, J. R., Macagno, F., Nespor, M., & Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences*, *109*(44), 17908-17913. https://doi.org/10.1073/pnas.1205413109

Benguerel, A. P., & Pichora-Fuller, M. K. (1982). Coarticulation effects in lipreading. *Journal of Speech, Language, and Hearing Research*, *25*(4), 600-607. https://doi.org/10.1044/jshr.2504.600

Berdasco-Muñoz, E., Nazzi, T., & Yeung, H. H. (2019). Visual scanning of a talking face in preterm and full-term infants. *Developmental Psychology, 55*(7), 1353 1361. https://doi.org/10.1037/dev0000737

Berg, T. (1999). Stress variation in British and American English. *World Englishes*, *18*(2), 123-143. https://doi.org/10.1111/1467-971X.00128

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 of the United States of America*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Bergelson, E., & Swingley, D. (2015). Early word comprehension in infants: Replication and extension. *Language Learning and Development*, *11*(4), 369-380. https://doi.org/10.1080/15475441.2014.979387

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, *117*(1), 21–33. https://doi.org/10.1037/0096-3445.117.1.21

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language* research. Timonium, MD: York Press.

Best, C. T., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*, *4*, 295. https://doi.org/10.1016/S0163-6383(98)91508-9

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development*, *18*(3), 339-350. https://doi.org/10.1016/0163-6383(95)90022-5

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345. https://doi:10.1037/0096-1523.14.3.345

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*(3), 305-330. https://doi.org/10.1016/S0095-4470(19)30637-0

Bijeljac-Babic, R., Serres, J., Höhle, B., & Nazzi, T. (2012). Effect of bilingualism on lexical stress pattern discrimination in French-learning infants. *PLoS One*, *7*(2), e30843. https://doi.org/10.1371/journal.pone.0030843

Birnholz, J. C., & Benacerraf, B. R. (1983). The development of human fetal hearing. *Science*, *222*(4623), 516-518.

Birulés, J., Bosch, L., Pons, F., & Lewkowicz, D. J. (2020). Highly proficient L2 speakers still need to attend to a talker's mouth when processing L2 speech. *Language, Cognition and Neuroscience*, *35*(10), 1314-1325. https://doi.org/10.1080/23273798.2020.1762905

Bleses, D., Basbøll, H., & Vach, W. (2011). Is Danish difficult to acquire? Evidence from Nordic past-tense studies. *Language and Cognitive Processes*, *26*(8), 1193-1231. https://doi.org/10.1080/01690965.2010.515107

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language, 35*(3), 619–650. https://doi.org/10.1017/S0305000908008714

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer. Retrieved from http://www.fon.hum.uva.nl/praat/.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*(6), 451–459. https://doi.org/10.1111/j.0956-7976.2005.01556.x

Bond, Z. S., & Garnes, S. (1980). Misperceptions of fluent speech. In R. A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science*, *16*(4), 298-304. https://doi.org/10.1111/j.0956-7976.2005.01531.x

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/10.1177/00238309030460020801

Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2015). Call me Alix, not Elix: vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, *18*(4), 587-598. https://doi.org/10.1111/desc.12242

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33-B44. https://doi.org/10.1016/S0010-0277(01)00122-6

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J. F. (2008). Hearing faces: how the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, *21*(5), 905-921. https://doi.org/10.1162/jocn.2009.21076

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535-543. https://doi.org/10.1111/j.1467-7687.2005.00445.x

Buchan, J. N., Paré, M., & Munhall, K. G. (2008). The effect of varying talker identity and listening conditions on gaze behavior during audiovisual speech perception. *Brain Research*, *1242*, 162-171. https://doi.org/10.1016/j.brainres.2008.06.083

Burnham, D., & Dodd, B. (2004). Auditory–visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*(4), 204-220. https://doi.org/10.1002/dev.20032

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28*(3), 455-474. https://doi.org/10.1017/S0142716407070257

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science*, *21*(3), 343-348. https://doi.org/10.1177/0956797609360758

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature*, *403*(6768), 428-430. https://doi.org/10.1038/35000206

Carreiras, M., Gillon-Dowens, M., Vergara, M., & Perea, M. (2008). Are vowels and consonants processed differently? Event-related potential evidence with a delayed letter paradigm. *Journal of Cognitive Neuroscience*, *21*(2), 275-288. https://doi.org/10.1162/jocn.2008.21023

Carreiras, M., & Price, C. J. (2008). Brain activation for consonants and vowels. *Cerebral Cortex*, *18*(7), 1727-1735. https://doi.org/10.1093/cercor/bhm202

Carreiras, M., Vergara, M., & Perea, M. (2007). ERP correlates of transposed-letter similarity effects: Are consonants processed differently from vowels?. *Neuroscience Letters*, *419*(3), 219-224. https://doi.org/10.1016/j.neulet.2007.04.053

Casillas, M., & Cristia, A. (2019). A step-by-step guide to collecting and analyzing long-format speech environment (LFSE) recordings. *Collabra: Psychology*, *5*(1), 24. https://doi.org/10.1525/collabra.209

Cassia, V. M., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference?. *Psychological Science*, *15*(6), 379–383. https://doi.org/10.1111/j.0956-7976.2004.00688.x

Chan, K. Y., & Vitevitch, M. S. (2015). The influence of neighborhood density on the recognition of Spanish-accented words. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(1), 69-85. https://doi.org/10.1037/a0038347

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436

Chen, H., Lee, D.T., Luo, Z., Lai, R.Y., Cheung, H., & Nazzi, T. (2021). Variation in phonological bias: Bias for vowels, rather than consonants or tones in lexical processing by Cantonese-learning toddlers. *Cognition*, 104486. https://doi.org/10.1016/j.cognition.2020.104486

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353. https://doi.org/10.1038/1561

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., & Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Research*, *82*(1), 53–58. https://doi.org/10.1016/0378-5955(94)00164-L

Christophe, A., Guasti, T., & Nespor, M. (1997). Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, *12*(5-6), 585-612. https://doi.org/10.1080/016909697386637

Christophe, A., Mehler, J., & Sebastian-Galles, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, *2*(3), 385–394. https://doi.org/10.1207/S15327078IN0203_6

Cole, R. A., & Scott, B. (1974). The phantom in the phoneme: Invariant cues for stop consonants. *Perception & Psychophysics*, *15*(1), 101-107. https://doi.org/10.3758/BF03205836

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*(1), 337-367. https://doi.org/10.1146/annurev.psych.52.1.337

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language*, *54*(1), 1-19. https://doi.org/10.1016/j.jml.2005.09.003

Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development*, *85*(4), 1330-1345. https://doi.org/10.1111/cdev.12193

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology, 52(4)*, 521–536. https://doi.org/10.1037/dev0000083

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word–object associative learning. *Developmental Science*, *12*(5), 725-731. https://doi.org/10.1111/j.1467-7687.2009.00814.x

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*, *2*(3-4), 133-142. https://doi.org/10.1016/0885-2308(87)90004-0

Cutler, A., & Otake, T. (2002). Rhythmic categories in spoken-word recognition. *Journal of Memory and Language*, *46*(2), 296-322.https://doi.org/10.1006/jmla.2001.2814

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition*, *28*, 746-755. https://doi.org/10.3758/BF03198409

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*, *42*, 37-48. https://doi.org/10.1016/j.cogdev.2017.02.004

DeBolt, M. C., & Oakes, L. M. (2023). The impact of face masks on infants' learning of faces: An eye tracking study. *Infancy*, *28*(1), 71-91. **https://doi.org/10.1111/infa.12516**

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174-1176. https://doi.org/10.1126/science.7375928

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*(2), 133-150. https://doi.org/10.1016/0163-6383(86)90025-1

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, *370*(6487), 292-295. https://doi.org/10.1038/370292a0

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, *298*(5600), 2013-2015. https://doi.org/10.1126/science.1077066

Delattre, P. (1964). Comparing the vocalic features of English, German, Spanish, and French. *International Review of Applied Linguistics in Language Teaching*, *2*, 71–97. https://doi.org/10.1515/iral.1964.2.1.71

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics in Language Teaching*, *4*(14), 183–198. https://doi.org/10.1515/iral.1966.4.1-4.183

Delle Luche, C., Floccia, C., Granjon, L., & Nazzi, T. (2017). Infants' first words are not phonetically specified: Own name recognition in British English-learning 5-month-olds. *Infancy*, *22*(3), 362-388. https://doi.org/10.1111/infa.12151

Delle Luche, C., Poltrock, S., Goslin, J., New, B., Floccia, C., & Nazzi, T. (2014). Differential processing of consonants and vowels in the auditory modality: A cross-linguistic study. *Journal of Memory and Language*, *72*, 1-15. https://doi.org/10.1016/j.jml.2013.12.001

Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants?. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *45*(4), 187-203. https://doi.org/10.1002/dev.20033

Di Giorgio, E., Turati, C., Altoè, G., & Simion, F. (2012). Face detection in complex visual displays: An eye-tracking study with 3-and 6-month-old infants and adults. *Journal of Experimental Child Psychology*, *113*(1), 66-77. https://doi.org/10.1016/j.jecp.2012.04.012

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, *9*(6), 719-721. https://doi.org/10.1068/p090719

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in-and out-of-synchrony. *Cognitive Psychology*, *11*(4), 478-484. https://doi.org/10.1016/0010-0285(79)90021-5

Dorn, K., Cauvet, E., & Weinert, S. (2021). A cross-linguistic study of multisensory perceptual narrowing in German and Swedish infants during the first year of life. *Infant and Child Development*, *30*(2), e2217. https://doi.org/10.1002/icd.2217

Duñabeitia, J. A., & Carreiras, M. (2011). The relative position priming effect depends on whether letters are vowels or consonants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1143-1163. https://doi.org/10.1037/a0023577

Durkin, K., Rutter, D. R., & Tucker, H. (1982). Social interaction and language acquisition: Motherese help you. *First Language*, *3*(8), 107-120. https://doi.org/10.1177/014272378200300803

Duta, M. D., Styles, S. J., & Plunkett, K. (2012). ERP correlates of unexpected word forms in a picture–word study of infants and adults. *Developmental Cognitive Neuroscience*, *2*(2), 223-234. https://doi.org/10.1016/j.dcn.2012.01.003

Eggermont, J. J., & Moore, J. K. (2012). Morphological and functional development of the auditory nervous system. In L. A. Werner, R. R. Fay, & A. N. Popper (Eds.), *Human auditory development*. New York, NY: Springer.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, *16*(3), 513–521. https://doi.org/10.3758/BF03198580

Eimas, P. D., & Miller, J. L. (1980). Discrimination of information for manner of articulation. *Infant Behavior & Development*, *3*(4), 367–375. https://doi.org/10.1016/S0163-6383(80)80044-0

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(3968), 303–306. https://doi.org/10.1126/science.171.3968.303

Escudero, P., Best, C. T., Kitamura, C., & Mulak, K. E. (2014). Magnitude of phonetic distinction predicts success at early word learning in native and non-native accents. *Frontiers in Psychology*, *5*, 1059. https://doi.org/10.3389/fpsyg.2014.01059

Escudero, P., Mulak, K. E., & Vlach, H. A. (2016). Cross-situational learning of minimal word pairs. *Cognitive Science, 40*(2), 455–465. https://doi.org/10.1111/cogs.12243

Faris, M. M., Best, C. T., & Tyler, M. D. (2018). Discrimination of uncategorised non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics*, *70*, 1-19. https://doi.org/10.1016/j.wocn.2018.05.003

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107. https://doi.org/10.1016/j.cognition.2016.03.005

Fava, E., Hull, R., & Bortfeld, H. (2014). Dissociating cortical activity during processing of native and non-native audiovisual speech from early to late infancy. *Brain Sciences*, *4*(3), 471-487. https://doi.org/10.1080/09297049.2013.803524

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *The Journal of the Acoustical Society of America*, *97*(3), 1893-1904. https://doi.org/10.1121/1.412063

Feld, J., & Sommers, M. (2011). There goes the neighborhood: Lipreading and the structure of the mental lexicon. *Speech Communication*, *53*(2), 220–228. https://doi.org/10.1016/j.specom.2010.09.003

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), v–173. https://doi.org/10.2307/1166093

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development*, *8*(2), 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development*, *10*(3), 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*(3), 477–501. https://doi.org/10.1017/S0305000900010679

Ferrand, L., & Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society*, *31*(2), 119-122. https://doi.org/10.3758/BF03334157

Ferreres, A. R., López, C. V., & China, N. N. (2003). Phonological alexia with vowel–consonant dissociation in non-word reading. *Brain and Language*, *84*(3), 399-413. https://doi.org/10.1016/S0093-934X(02)00559-X

Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*(1), 67–83. https://doi.org/10.1016/S0095-4470(19)30538-8

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J., & Vihman, M. (2016). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition, 148,* 1–9. https://doi.org/10.1016/j.cognition.2015.12.004

Floccia, C., Nazzi, T., Delle Luche, C., Poltrock, S., & Goslin, J. (2014). English-learning one-to two-year-olds do not show a consonant bias in word learning. *Journal of Child Language*, *41*(5), 1085-1114. https://doi:10.1017/S0305000913000287

Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot-Smith, K., Krott, A., Mills, D., Rowland, C., Gervain, J., & Plunkett, K. (2018). Vocabulary of 2-year-olds learning English and an additional language: Norms and effects of linguistic distance. *Monographs of the Society for Research in Child Development*, *83*(1), 7–29. https://doi.org/10.1111/mono.12348

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 680–698. https://doi.org/10.1037/0278-7393.10.4.680

Fort, M., Kandel, S., Chipot, J., Savariaux, C., Granjon, L., & Spinelli, E. (2013). Seeing the initial articulatory gestures of a word triggers lexical access. *Language and Cognitive Processes*, *28*(8), 1207–1223. https://doi.org/10.1080/01690965.2012.701758

Fort, M., Spinelli, E., Savariaux, C., & Kandel, S. (2010). The word superiority effect in audiovisual speech perception. *Speech Communication*, *52*(6), 525–532. https://doi.org/10.1016/j.specom.2010.02.005

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677-694. https://doi.org/10.1017/S0305000916000209

Fraser, S., Gagne´, J. P., Alepins, M., & Dubois, P. (2010). Evaluating the effort expended to understand speech in noise using a dual-task paradigm: The effects of providing visual speech cues. *Journal of Speech, Language, and Hearing Research*, *53*(1), 18–33. https://doi.org/10.1044/1092-4388(2009/08-0140)

Friedrich, M., & Friederici, A. D. (2017). The origins of word learning: Brain responses of 3-month-olds indicate their rapid association of objects and words. *Developmental Science*, *20*(2), e12357. https://doi.org/10.1111/desc.12357

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, *5*(4), 171-189. https://doi.org/10.1177/002383096200500401

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*(2), 149– 175. https://doi.org/10.1016/S0095-4470(19)30813-7

Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, *4*(1), 1490. https://doi.org/10.1038/ncomms2430

Giegrich, H. J. (1992). *English phonology. An introduction*. Cambridge, UK: Cambridge University Press.

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Oller, D. K., Hansen, J. H. L., & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology*, *26*(2), 248–265. https://doi.org/10.1044/2016_AJSLP-15-0169

Gómez, D. M., Mok, P., Ordin, M., Mehler, J., & Nespor, M. (2018). Statistical speech segmentation in tone languages: The role of lexical tones. *Language and Speech*, *61*(1), 84-96. https://doi.org/10.1177/0023830917706529

Gonzalez-Gomez, N., & Nazzi, T. (2013). Effects of prior phonotactic knowledge on infant word segmentation: the case of nonadjacent dependencies. *Journal of Speech, Language, and Hearing* Research, 56, 840-849. https://doi.org/10.1044/1092-4388(2012/12-0138)

Goyet, L., de Schonen, S., & Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. *Brain Research*, *1332*, 75-89. https://doi.org/10.1016/j.brainres.2010.03.047

Goyet, L., Nishibayashi, L. L., & Nazzi, T. (2013). Early syllabic segmentation of fluent speech by infants acquiring French. *PloS One*, *8*(11), e79646. https://doi.org/10.1371/journal.pone.0079646

Grainger, J., Granier, J.P., Farioli, F., Van Assche, E., & van Heuven, W. J. B. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 865–884. https://doi.org/10.1037/0096-1523.32.4.865

Granier-Deferre, C., Ribeiro, A., Jacquet, A. Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science*, *14*(2), 336-352. https://doi.org/10.1111/j.1467-7687.2010.00978.x

Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, *108*(3), 1197-1208. https://doi.org/10.1121/1.1288668

Grant, K. W., Van Wassenhove, V., & Poeppel, D. (2004). Detection of auditory (cross-spectral) and auditory–visual (cross-modal) synchrony. *Speech Communication*, *44*(1-4), 43-53. https://doi.org/10.1016/j.specom.2004.06.004

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, *25*(4), 577–588. https://doi.org/10.1037/0012-1649.25.4.577

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*(5), 2711–2724. https://doi.org/10.1121/1.428657

Gurler, D., Doyle, N., Walker, E., Magnotti, J., & Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. *Attention, Perception, & Psychophysics*, *77*(4), 1333–1341. https://doi.org/10.3758/s13414-014-0821-1

Hallé, P. A., & de Boysson-Bardies, B. (1994). Emergence of an early receptive lexicon: Infants' recognition of words. *Infant Behavior and Development*, *17*(2), 119-129. https://doi.org/10.1016/0163-6383(94)90047-7

Hallé, P. A., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, *19*(4), 463-481. https://doi.org/10.1016/S0163-6383(96)90007-7

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British communicative development inventory. *Journal of Child Language*, *27*(3), 689-705. https://doi.org/10.1017/S0305000900004414

Havy, M., & Nazzi, T. (2009). Better processing of consonantal over vocalic information in word learning at 16 months of age. *Infancy*, *14*(4), 439–456. https://doi.org/10.1080/15250000902996532

Havy, M., Serres, J., & Nazzi, T. (2014). A consonant/vowel asymmetry in word-form processing: Evidence in childhood and in adulthood. *Language and Speech*, *57*(2), 254–281. https://doi.org/10.1177/0023830913507693

Hawthorne, K., & Gerken, L. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition*, *133*(2), 420-428. https://doi.org/10.1016/j.cognition.2014.07.013

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*(3), 360-378. https://doi.org/10.1016/j.specom.2005.04.007

Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*(2), 298–310. https://doi.org/10.1044/1092-4388(2009/08-0243)

Hirst, D., Di Cristo, A., & Nishinuma, Y. (2001). Prosodic parameters of French: A cross-language approach. In *Contrastive studies of Japanese and other languages series*. Tokyo: National Institute for Japanese Language.

Hochmann, J.R., Benavides-Varela, S., Fló, S., Nespor, M., & Mehler, J. (2017). Bias for vocalic over consonantal information in 6-month-olds. Infancy, 23(1), 136–151. https://doi.org/10.1111/infa.12203

Hochmann, J. R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: different roles in early language acquisition. *Developmental Science*, *14*(6), 1445-1458. https://doi.org/10.1111/j.1467-7687.2011.01089.x

Højen, A., & Nazzi, T. (2016). Vowel bias in Danish word-learning: processing biases are language-specific. *Developmental Science*, *19*(1), 41-49. https://doi.org/10.1111/desc.12286

Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, *76*(3), 598–613. https://doi.org/10.1111/j.1467-8624.2005.00866.x

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research, 5,* 69–95.

Imafuku, M., Kanakogi, Y., Butler, D., & Myowa, M. (2019). Demystifying infant vocal imitation: The roles of mouth looking and speaker's gaze. *Developmental Science*, *22*(6), e12825. https://doi.org/10.1111/desc.12825

Issard, C., Tsuji, S., & Cristia, A. (2021). Infants' preference for speech is stable across the first year of life: Meta-analytic evidence. *Infancy*, 1– 19. https://doi.org/10.1111/infa.12529

Jansson-Verkasalo, E., Ruusuvirta, T., Huotilainen, M., Alku, P., Kushnerenko, E., Suominen, K., Rytky, S., Luotonen, M., Kaukola, T., Tolonen, U., & Hallman, M. (2010). Atypical perceptual narrowing in prematurely born infants is associated with compromised language acquisition at 2 years of age. *BMC Neuroscience*, *11*, 88. https://doi.org/10.1186/1471-2202-11-88

JASP Team (2022). JASP (Version 0.16.3) [Computer software].

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567. https://doi.org/10.1006/jmla.2000.2755

Johnson, E. K., Jusczyk, P. W., Cutler, A., & Norris, D. (2003). Lexical viability constraints on speech segmentation by infants. *Cognitive Psychology*, *46*(1), 65–97. https://doi.org/10.1016/S0010-0285(02)00507-8

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1-2), 1–19. https://doi.org/10.1016/0010-0277(91)90045-6

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. https://doi.org/10.1006/cogp.1995.1010

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675-687. https://doi.org/10.1111/j.1467-8624.1993.tb02935.x

Jusczyk, P. W., Goodman, M. B., & Baumann, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language, 40*(1), 62–82. https://doi.org/10.1006/jmla.1998.2605

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*(3-4), 159–207. https://doi.org/10.1006/cogp.1999.0716

Kaganovich, N., Schumaker, J., & Rowland, C. (2016). Matching heard and seen speech: an ERP study of audiovisual word recognition. *Brain and Language*, *157*, 14-24. https://doi.org/10.1016/j.bandl.2016.04.010

Keidel, J. L., Jenison, R. L., Kluender, K. R., & Seidenberg, M. S. (2007). Does grammar constrain statistical learning? Commentary on Bonatti, Peña, Nespor, and Mehler (2005). *Psychological Science*, *18*(10), 922-923. https://doi.org/10.1111/j.1467-9280.2007.02001.x

Kern, S., & Géraldine, H. (2003). Development of communicative gestures in French infants from 8 to 16 months. *Citeseer*, 1–14.

Kim, J., Davis, C., & Krins, P. (2004). Amodal processing of visual speech as revealed by priming. *Cognition*, *93*(1), B39–B47. https://doi.org/10.1016/j.cognition.2003.11.003

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*(3), 279–312. https://doi.org/10.1016/S0095-4470(19)31059-9

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Cognitive Brain Research*, *24*(1), 109–116. https://doi.org/10.1016/j.cogbrainres.2004.12.009

Kubicek, C., Gervain, J., De Boisferon, A. H., Pascalis, O., Lœvenbruck, H., & Schwarzer, G. (2014). The influence of infant-directed speech on 12-month-olds' intersensory perception of fluent speech. *Infant Behavior and Development*, *37*(4), 644-651. https://doi.org/10.1016/j.infbeh.2014.08.010

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107. https://doi.org/10.3758/BF03212211

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The Native Language Magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life*. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-015-8234-6_22

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979-1000. https://doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*(4577), 1138–1141. https://doi.org/10.1126/science.7146899

Kuhl, P. K., & Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics*, *31*(3), 279–292. https://doi.org/10.3758/BF03202536

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608. https://doi.org/10.1126/science.1736364

Kujala, A., Huotilainen, M., Hotakainen, M., Lennes, M., Parkkonen, L., Fellman, V., & Näätänen, R. (2004). Speech-sound discrimination in neonates as measured with MEG. *NeuroReport: For Rapid Communication of Neuroscience Research*, *15*(13), 2089–2092. https://doi.org/10.1097/00001756-200409150-00018

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11442–11445. https://doi.org/10.1073/pnas.0804275105

Kushnerenko, E., Tomalski, P., Ballieux, H., Potton, A., Birtles, D., Frostick, C., & Moore, D. G. (2013). Brain responses and looking behavior during audiovisual speech integration in infants predict auditory speech comprehension in the second year of life. *Frontiers in Psychology*, *4*, 432. https://doi.org/10.3389/fpsyg.2013.00432

Ladefoged, P., & Disner, S.F. (2012). *Vowels and Consonants* (3rd Ed.). Chichester, UK: Wiley-Blackwell.

Ladefoged, P., & Johnson, K. (2015). *A Course in Phonetics* (7th Ed.). Boston, MA: Wadsworth, Cengage.

Lalonde, K., & Holt, R. F. (2015). Preschoolers benefit from visually salient speech cues. *Journal of Speech, Language, and Hearing Research*, *58*(1), 135-150. https://doi.org/10.1044/2014_JSLHR-H-13-0343

Lalonde, K., & Werner, L. A. (2019). Infants and adults use visual cues to improve detection and discrimination of speech in noise. *Journal of Speech, Language, and Hearing Research*, *62*(10), 3860-3875. https://doi.org/10.1044/2019_JSLHR-H-19-0106

Lalonde, K., & Werner, L. A. (2021). Development of the mechanisms underlying audiovisual speech perception benefit. *Brain Sciences*, *11*(1), 49. https://doi.org/10.3390/brainsci11010049

Lecanuet, J. P., & Schaal, B. (2002). Sensory performances in the human foetus: A brief summary of research. *Intellectica*, *34*(1), 29-56.

Leppänen, J. M. (2016). Using eye tracking to understand infants' attentional bias for faces. *Child Development Perspectives*, *10*(3), 161–165. https://doi.org/10.1111/cdep.12180

Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, *14*(6), 1323–1329. https://doi.org/10.1111/j.1467-7687.2011.01079.x

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109

Lewkowicz, D. J., & Pons, F. (2013). Recognition of amodal language identity emerges in infancy. *International Journal of Behavioral Development*, *37*(2), 90–94. https://doi.org/10.1177/0165025412467582

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. https://doi.org/10.1037/h0020279

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384-422. https://doi.org/10.1080/00437956.1964.11659830

Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, *133*(2), 385–394. https://doi.org/10.1016/j.cognition.2014.06.004

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1-36. https://doi.org/10.1097/00003446-199802000-00001,

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science*, *219*(4590), 1347–1349. https://doi.org/10.1126/science.6828865

Maddieson, I. (1984). *Patterns of sounds*. Cambridge, UK: Cambridge University Press.

Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., Grebe, R., & Wallois, F. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proceedings of the National Academy of Sciences*, *110*(12), 4846-4851. https://doi.org/10.1073/pnas.1212220110

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, *19*(23), 1994-1997. https://doi.org/10.1016/j.cub.2009.09.064

Mandel, D. R., Jusczyk, P. W., & Pisoni, D. B. (1995). Infants' recognition of the sound patterns of their own names. *Psychological Science*, *6*(5), 314–317. https://doi.org/10.1111/j.1467-9280.1995.tb00517.x

Mani, N., Mills, D. L., & Plunkett, K. (2012). Vowels in early words: An event-related potential study. *Developmental Science*, *15*(1), 2-11. https://doi.org/10.1111/j.1467-7687.2011.01092.x

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, *57*(2), 252–272. https://doi.org/10.1016/j.jml.2007.03.005

Mani, N., & Plunkett, K. (2010). Twelve-month-olds know their cups from their keps and tups. *Infancy*, *15*(5), 445-470. https://doi.org/10.1111/j.1532-7078.2009.00027.x

Männel, C., & Friederici, A. D. (2013). Accentuate or repeat? Brain signatures of developmental periods in infant word recognition. *Cortex*, *49*(10), 2788-2798. https://doi.org/10.1016/j.cortex.2013.09.003

Marks, E. A., Moates, D. R., Bond, Z. S., & Vasquez, L. (2002). Vowel mutability: The case of monolingual Spanish listeners and bilingual Spanish-English listeners. *Southwest Journal of Linguistics*, *21*(2), 73-100.

Mattys, S. L., Bernstein, L. E., & Auer, E. T. (2002). Stimulus-based lexical distinctiveness as a general word-recognition mechanism. *Perception & Psychophysics*, *64*(4), 667–679. https://doi.org/10.3758/BF03194734

Mattys, S. L., & Jusczyk, P. W. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 644–655. https://doi.org/10.1037/0096-1523.27.3.644

May, L., Byers-Heinlein, K., Gervain, J., & Werker, J. F. (2011). Language and the newborn brain: Does prenatal language experience shape the neonate neural response to speech? *Frontiers in Psychology*, *2*, 222. https://doi.org/10.3389/fpsyg.2011.00222

May, L., Gervain, J., Carreiras, M., & Werker, J. F. (2018). The specificity of the neural response to speech at birth. *Developmental Science*, *21*(3), e12564. https://doi.org/10.1111/desc.12564

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134. https://doi.org/10.1111/j.1467-7687.2007.00653.x

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111. https://doi.org/10.1016/S0010-0277(01)00157-3

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from CDI analysis. *Developmental Science*, *14*(4), 769-785. https://doi.org/10.1111/j.1467-7687.2010.01024.x

McCarthy, J. (1985), *Formal problems in Semitic phonology and morphology*, New York, NY: Garland.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748. https://doi.org/10.1038/264746a0

McQueen, J. (1996). Phonetic categorisation. *Language and Cognitive Processes*, *11*(6), 655-664. https://doi.org/10.1080/016909696387060

Mehler, J., Bertoncini, J., Barrière, M., & Jassik-Gerschenfeld, D. (1978). Infant recognition of mother's voice. *Perception*, *7*(5), 491–497. https://doi.org/10.1068/p070491

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

Mehler, J. & Nespor, M. (2003). Linguistic rhythm and the development of language. In A Belletti & L. Rizzi (Eds.), *Structures and beyond: The cartography of syntactic structures*. Oxford, UK: Oxford University Press.

Mehler, J., Peña, M., Nespor, M., & Bonatti, L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex*, *42*(6), 846–854. https://doi.org/10.1016/S0010-9452(08)70427-1

Mercure, E., Bright, P., Quiroz, I., & Filippi, R. (2022). Effect of infant bilingualism on audiovisual integration in a McGurk task. *Journal of Experimental Child Psychology*, *217*, 105351. https://doi.org/10.1016/j.jecp.2021.105351

Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M. H., & MacSweeney, M. (2019). Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Developmental Science*, *22*(1), e12701. https://doi.org/10.1111/desc.12701

Mersad, K., & Nazzi, T. (2012). When mommy comes to the rescue of statistics: Infants combine top-down and bottom-up cues to segment speech. *Language Learning and Development*, *8*(3), 303–315. https://doi.org/10.1080/15475441.2011.609106

Miller, G. A., & Nicely, P. A. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, *27*, 338–352. https://doi.org/10.1121/1.1907526

Mitchel, A. D., & Weiss, D. J. (2014). Visual speech segmentation: Using facial cues to locate word boundaries in continuous speech. *Language, Cognition and Neuroscience*, *29*(7), 771–780. https://doi.org/10.1080/01690965.2013.791703

Monte-Ordoño, J., & Toro, J. M. (2017). Different ERP profiles for learning rules over consonants and vowels. *Neuropsychologia*, *97*, 104–111. https://doi.org/10.1016/j.neuropsychologia.2017.02.014

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior & Development*, *16*(4), 495–500. https://doi.org/10.1016/0163-6383(93)80007-U

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced *in utero* affects vowel perception after birth: A two-country study. *Acta Paediatrica*, *102*(2), 156–160. https://doi.org/10.1111/apa.12098

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior & Development*, *21*(2), 373–377. https://doi.org/10.1016/S0163-6383(98)90014-5

Morgan, J. L. (1996). Prosody and the roots of parsing. *Language and Cognitive Processes*, *11*(1-2), 69–106. https://doi.org/10.1080/016909696387222

Morgan, J. L., & Demuth, k. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Morikawa, H., Shand, N., & Kosawa, Y. (1988). Maternal speech to prelingual infants in Japan and the United States: Relationships among functions, forms and referents. *Journal of Child Language*, *15*(2), 237–256. https://doi.org/10.1017/S0305000900012356

Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Byers-Heinlein, K. (2019). Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Developmental Psychology*, *55*(8), 1640–1655. https://doi.org/10.1037/dev0000750

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407–420. https://doi.org/10.1111/j.1467-7687.2009.00898.x

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, *98*(1), 13-30. https://doi.org/10.1016/j.cognition.2004.10.005

Nazzi, T., & Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: New evidence from consonants in coda positions. *Language and Speech*, *52*(4), 463-480. https://doi.org/10.1177/0023830909336584

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 756–766. https://doi.org/10.1037/0096-1523.24.3.756

Nazzi, T., & Cutler, A. (2019). How consonants and vowels shape spoken-language recognition. *Annual Review of Linguistics*, *5*, 25-47. https://doi.org/10.1146/annurev-linguistics-011718-011919

Nazzi, T., Floccia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology*, *102*(4), 522-537. https://doi.org/10.1016/j.jecp.2008.05.003

Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: when does language become a tool for categorization?. *Cognition*, *80*(3), B11-B20. https://doi.org/10.1016/S0010-0277(01)00112-3

Nazzi, T., & New, B. (2007). Beyond stop consonants: Consonantal specificity in early lexical acquisition. *Cognitive Development*, *22*(2), 271-279. https://doi.org/10.1016/j.cogdev.2006.10.007

Nazzi, T., & Polka, L. (2018). The consonant bias in word learning is not determined by position within the word: Evidence from vowel-initial words. *Journal of Experimental Child Psychology*, *174*, 103–111. https://doi.org/10.1016/j.jecp.2018.05.011

Nazzi, T., Poltrock, S., & Von Holzen, K. (2016). The developmental origins of the consonant bias in lexical processing. *Current Directions in Psychological Science*, *25*(4), 291-296. https://doi.org/10.1177/0963721416655786

Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior & Development*, *18*(1), 111–116. https://doi.org/10.1016/0163-6383(95)90012-8

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *2*(2), 203-230. https://doi.org/10.1418/10879

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht, Netherlands: Foris Publications.

New, B., Araújo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological Science*, *19*(12), 1223–1227. https://doi.org/10.1111/j.1467-9280.2008.02228.x

New, B., & Nazzi, T. (2014). The time course of consonant and vowel processing during word recognition. *Language, Cognition and Neuroscience*, *29*(2), 147–157. https://doi.org/10.1080/01690965.2012.735678

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments & Computers*, *36*(3), 516–524. https://doi.org/10.3758/BF03195598

Newport, E., Gleitman, H. & Gleitman, L. (1977). Mother, I'd rather do it myself: some effects and non-effects of maternal speech style. In Snow, C. & Ferguson, C. (Eds.), *Talking to children: language input and acquisition*. Cambridge, UK: Cambridge University Press.

Nishibayashi, L. L., Goyet, L., & Nazzi, T. (2015). Early speech segmentation in French-learning infants: Monosyllabic words versus embedded syllables. *Language and Speech*, *58*(3), 334–350. https://doi.org/10.1177/0023830914551375

Nishibayashi, L. L., & Nazzi, T. (2016). Vowels, then consonants: Early bias switch in recognizing segmented word forms. *Cognition*, *155*, 188-203. https://doi.org/10.1016/j.cognition.2016.07.003

Nomikou, I., Rohlfing, K. J., Cimiano, P., & Mandler, J. M. (2019). Evidence for early comprehension of action verbs. *Language Learning and Development*, *15*(1), 64-74. https://doi.org/10.1080/15475441.2018.1520639

Oller, D. K. (2000). *The emergence of the speech capacity.* Mahwah, NJ: Erlbaum.

Owens, E., & Blazek, B. (1985). Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech & Hearing Research*, *28*(3), 381–393. https://doi.org/10.1044/jshr.2803.381

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), B9–B17. https://doi.org/10.1016/S0010-0277(97)00030-9

Palmer, S. B., Fais, L., Golinkoff, R. M., & Werker, J. F. (2012). Perceptual narrowing of linguistic sign occurs in the 1st year of life. *Child Development*, *83*(2), 543–553. https://doi.org/10.1111/j.1467-8624.2011.01715.x

Paré, M., Richler, R. C., ten Hove, M., & Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception & Psychophysics*, *65*(4), 553–567. https://doi.org/10.3758/BF03194582

Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, *23*(7), 728–733. https://doi.org/10.1177/0956797612438734

Parise, E., Friederici, A. D., & Striano, T. (2010). “Did you call me?” 5-month-old infants own name guides their attention. *PLoS One*, *5*(12), e14208. https://doi.org/10.1371/journal.pone.0014208

Partanen, E., Kujala, T., Tervaniemi, M., & Huotilainen, M. (2013). Prenatal music exposure induces long-term neural effects. *PloS one*, *8*(10), e78946. https://doi.org/10.1371/journal.pone.0078946

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language*, 384-402.

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191–196. https://doi.org/10.1111/1467-7687.00271

Peelle, J. E., & Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, *68*, 169–181. https://doi.org/10.1016/j.cortex.2015.03.006

Pejovic, J., Yee, E., & Molnar, M. (2020). Speaker matters: Natural inter-speaker variation affects 4-month-olds' perception of audio-visual speech. *First Language*, *40*(2), 113-127. https://doi.org/10.1177/0142723719876382

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, *100*(20), 11702-11705. https://doi.org/10.1073/pnas.1934290100

Peña, M., Werker, J. F., & Dehaene-Lambertz, G. (2012). Earlier speech exposure does not accelerate speech acquisition. *The Journal of Neuroscience*, *32*(33), 11159–11163. https://doi.org/10.1523/JNEUROSCI.6516-11.2012

Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G., & Friederici, A. D. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences*, *108*(38), 16056-16061. https://doi.org/10.1073/pnas.1102991108

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO! Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, *51*(2), 231–246. https://doi.org/10.1016/j.jml.2004.05.005

Pickens, J., Field, T., Nawrocki, T., Martinez, A., Soutullo, D., & Gonzalez, J. (1994). Full-term and preterm infants' perception of face-voice synchrony. *Infant Behavior & Development*, *17*(4), 447–455. https://doi.org/10.1016/0163-6383(94)90036-1

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*(2), 253–260. https://doi.org/10.3758/BF03214136

Polka, L., & Bohn, O.S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, *100*(1), 577–592. https://doi.org/10.1121/1.415884

Polka, L., & Bohn, O. S. (2011). Natural referent vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467 478. https://doi.org/10.1016/j.wocn.2010.08.007

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/-/ð/ perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America*, *109*(5), 2190–2201. https://doi.org/10.1121/1.1362689

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421-435. https://doi.org/10.1037/0096-1523.20.2.421

Poltrock, S., & Nazzi, T. (2015). Consonant/vowel asymmetry in early word form recognition. *Journal of Experimental Child Psychology*, *131*, 135-148. https://doi.org/10.1016/j.jecp.2014.11.011

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

Pons, F., Bosch, L., & Lewkowicz, D. J. (2019). Twelve-month-old infants’ attention to the eyes of a talking face is associated with communication and social skills. *Infant Behavior & Development*, *54*, 80–84. https://doi.org/10.1016/j.infbeh.2018.12.003

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10598–10602. https://doi.org/10.1073/pnas.0904134106

Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development*, *77*(2), 266–280. https://doi.org/10.1111/j.1467-8624.2006.00869.x

Quaresima, V., Bisconti, S., & Ferrari, M. (2012). A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain and Language*, *121*(2), 79-89. https://doi.org/10.1016/j.bandl.2011.03.009

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, *288*(5464), 349–351. https://doi.org/10.1126/science.288.5464.349

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Repp, B.H. (1984). Categorical perception: issues, methods, findings. In N.J. Lass (Ed.), *Speech and language: Advances in basic research and practice, Vol. 10*. New York, NY: Academic Press.

Reynolds, G. D., & Romano, A. C. (2016). The development of attention systems and working memory in infancy. *Frontiers in Systems Neuroscience*, *10*, 15. https://doi.org/10.3389/fnsys.2016.00015

Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual's gaze. *Neuroreport*, *13*(17), 2259–2264. https://doi.org/10.1097/00001756-200212030-00018

Ridouane, R. (2008). Syllables without vowels: phonetic and phonological evidence from Tashlhiyt Berber. *Phonology*, *25*(2), 321-359. https://doi.org/10.1017/S0952675708001498

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, *8*(2), 162–172. https://doi.org/10.1111/j.1467-7687.2005.00403.x

Roach, P. (2004). British English: Received pronunciation. *Journal of the International Phonetic Association*, *34*(2), 239-245. https://doi.org/10.1017/S0025100304001768

Robert-Ribes, J., Schwartz, J. L., Lallouache, T., & Escudier, P. (1998). Complementarity and synergy in bimodal speech: Auditory, visual, and audio-visual identification of French oral vowels in noise. *The Journal of the Acoustical Society of America*, *103*(6), 3677-3689. https://doi.org/10.1121/1.423069

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*(6), 405–409. https://doi.org/10.1111/j.1467-8721.2008.00615.x

Rosenblum, L. D., & Saldaña, H. M. (1992). Discrimination tests of visually influenced syllables. *Perception & Psychophysics*, *52*(4), 461–473. https://doi.org/10.3758/BF03206706

Rosenblum, L. D., & Saldaña, H. M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 318–331. https://doi.org/10.1037/0096-1523.22.2.318

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, *59*(3), 347–357. https://doi.org/10.3758/BF03211902

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147–1153. https://doi.org/10.1093/cercor/bhl024

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257. https://doi.org/10.1016/j.tics.2007.04.004

Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1996). Differential activation of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain*, *119*(6), 2073–2083. https://doi.org/10.1093/brain/119.6.2073

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, *45*(3), 587- 597. https://doi.org/10.1016/j.neuropsychologia.2006.03.036

Sato, H., Hirabayashi, Y., Tsubokura, H., Kanai, M., Ashida, T., Konishi, I., Uchida-Ota., M., Konishi, Y., & Maki, A. (2012). Cerebral hemodynamics in newborn infants exposed to speech sounds: A whole-head optical topography study. *Human Brain Mapping*, *33*(9), 2092-2103. https://doi.org/10.1002/hbm.21350

Schwartz, J. L., & Savariaux, C. (2014). No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag. *PLoS Computational Biology*, *10*(7), e1003743. https://doi.org/10.1371/journal.pcbi.1003743

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: Is the distributional account all there is to it?. *Developmental Science*, *12*(6), 874–887. https://doi.org/10.1111/j.1467-7687.2009.00829.x

Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science*, *11*(2), 306-320. https://doi.org/10.1111/j.1467-7687.2008.00677.x

Sekiyama, K., Hisanaga, S., & Mugitani, R. (2021). Selective attention to the mouth of a talker in Japanese-learning infants and toddlers: Its relationship with vocabulary and compensation for noise. *Cortex*, *140*, 145–156. https://doi.org/10.1016/j.cortex.2021.03.023

Shallice, T., Fletcher, P., Frith, C. D., Grasby, P., Frackowiak, R. S. J., & Dolan, R. J. (1994). Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature*, *368*(6472), 633–635. https://doi.org/10.1038/368633a0

Sharp, D. J., Scott, S. K., Cutler, A., & Wise, R. J. S. (2005). Lexical retrieval constrained by sound structure: The role of the left inferior frontal gyrus. *Brain and Language*, *92*(3), 309–319. https://doi.org/10.1016/j.bandl.2004.07.002

Shepherd, J., Davies, G., & Ellis, H. (1981). Studies of cue saliency. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces*. New York, NY: Academic Press.

Shultz, S., & Vouloumanos, A. (2010). Three-month-olds prefer speech to other naturally occurring signals. *Language Learning and Development*, *6*(4), 241–257. https://doi.org/10.1080/15475440903507830

Shultz, S., Vouloumanos, A., Bennett, R. H., & Pelphrey, K. (2014). Neural specialization for speech in the first months of life. *Developmental Science*, *17*(5), 766-774. https://doi.org/10.1111/desc.12151

Singh, L., Goh, H. H., & Wewalaarachchi, T. D. (2015). Spoken word recognition in early childhood: Comparative effects of vowel, consonant and lexical tone variation. *Cognition*, *142*, 1-11. https://doi.org/10.1016/j.cognition.2015.05.010

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, *17*(10), 2387–2399. https://doi.org/10.1093/cercor/bhl147

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., Limissuri, R. A., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, *12*(6), 914–919. https://doi.org/10.1111/j.1467-7687.2009.00835.x

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

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, *64*(5), 1358-1368. https://doi.org/10.1121/1.382102

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the Study of Speech*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Stevens, K. N., & House, A. S. (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America*, *27*, 484–493. https://doi.org/10.1121/1.1907943

Streeter, L. A. (1976). Language perception of 2-mo-old infants shows effects of both innate mechanisms and experience. *Nature*, *259*(5538), 39–41. https://doi.org/10.1038/259039a0

Sugden, N. A., Mohamed-Ali, M. I., & Moulson, M. C. (2014). I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective. *Developmental Psychobiology*, *56*(2), 249–261. https://doi.org/10.1002/dev.21183

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, *26*, 212–215. https://doi.org/10.1121/1.1907309

Summerfield, Q. (1987). Speech perception in normal and impaired hearing. *British Medical Bulletin*, *43*(4), 909-925. https://doi.org/10.1093/oxfordjournals.bmb.a072225

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

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-/ in monolingual and bilingual acquisition of English. *Cognition*, *100*(2), 369–388. https://doi.org/10.1016/j.cognition.2005.04.007

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, *8*(5), 432-443. https://doi.org/10.1111/j.1467-7687.2005.00432.x

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3617-3632. https://doi.org/10.1098/rstb.2009.0107

Swoboda, P. J., Morese, P. A., & Leavitt, L. A. (1976). Continuous vowel discrimination in normal and at risk infants. *Child Development*, *47*(2), 459–465. https://doi.org/10.2307/1128802

Syrnyk, C., & Meints, K. (2017). Bye-bye mummy–Word comprehension in 9-month-old infants. *British Journal of Developmental Psychology*, *35*(2), 202-217. https://doi.org/10.1111/bjdp.12157

Taga, G., Homae, F., & Watanabe, H. (2007). Effects of source-detector distance of near infrared spectroscopy on the measurement of the cortical hemodynamic response in infants. *Neuroimage*, *38*(3), 452-460.https://doi.org/10.1016/j.neuroimage.2007.07.050

Tan, S. H. J., Kalashnikova, M., & Burnham, D. (2023). Seeing a talking face matters: Infants' segmentation of continuous auditory-visual speech. *Infancy*, *28*(2), 277-300. https://doi.org/10.1111/infa.12509

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850–855. https://doi.org/10.1016/j.cognition.2008.05.009

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., Morgan, J. L., & Shah, R. J. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173–1190. https://doi.org/10.1017/S0305000914000725

Ter Schure, S., Junge, C., & Boersma, P. (2016). Discriminating non-native vowels on the basis of multimodal, auditory or visual information: Effects on infants' looking patterns and discrimination. *Frontiers in Psychology*, *7*, 525. https://doi.org/10.3389/fpsyg.2016.00525

Thierry, G., Vihman, M., & Roberts, M. (2003). Familiar words capture the attention of 11-month-olds in less than 250 ms. *Neuroreport*, *14*(18), 2307-2310. https://doi.org/10.1097/00001756-200312190-00004

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*(2), 172-175. https://doi.org/10.1111/j.1467-7687.2005.00432.x

Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, *17*(4), 432–444. https://doi.org/10.1111/j.1532-7078.2011.00084.x

Tomalski, P., Ribeiro, H., Ballieux, H., Axelsson, E. L., Murphy, E., Moore, D. G., & Kushnerenko, E. (2013). Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. *European Journal of Developmental Psychology*, *10*(5), 611-624. https://doi.org/10.1080/17405629.2012.728076

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science*, *19*(2), 137–144. https://doi.org/10.1111/j.1467-9280.2008.02059.x

Trehub, S. E. (1973). Infants' sensitivity to vowel and tonal contrasts. *Developmental Psychology*, *9*(1), 91–96. https://doi.org/10.1037/h0034999

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development*, *47*(2), 466–472. https://doi.org/10.2307/1128803

Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology*, *169*, 93–109. https://doi.org/10.1016/j.jecp.2018.01.002

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, *56*(2), 179–191. https://doi.org/10.1002/dev.21179

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. T. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: language-specific developmental changes in speech perception abilities. In *Conference of the International Speech Communication Association*. https://doi.org/10.21437/icslp.1994-438

Twomey, K. E., & Westermann, G. (2018). Learned labels shape pre-speech infants' object representations. *Infancy*, *23*(1), 61–73. https://doi.org/10.1111/infa.12201

Tye-Murray, N., Sommers, M., & Spehar, B. (2007). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification*, *11*(4), 233-241. https://doi.org/10.1177/1084713807307409

Tye-Murray, N., Spehar, B., Myerson, J., Sommers, M. S., & Hale, S. (2011). Crossmodal enhancement of speech detection in young and older adults: Does signal content matter?. *Ear and Hearing*, *32*(5), 650-655. https://doi.org/10.1097/AUD.0b013e31821a4578

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, *71*(1), 4-21. https://doi.org/10.1159/000356237

Van de Weijer, J. (1998). *Language input for word discovery*. (PhD Dissertation). MPI Series in Psycholinguistics 9, Radboud University Nijmegen. https://doi.org/10.17617/2.2057670

Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, *24*, 573-583. https://doi.org/10.3758/BF03201084

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102*(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, B. (2016). HomeBank: An online repository of daylong child-centered audio recordings. *Seminars in Speech and Language*, *37*(2), 128–141. https://doi.org/10.1055/s-0036-1580745

Vannasing, P., Florea, O., González-Frankenberger, B., Tremblay, J., Paquette, N., Safi, D., Wallois, F., Lepore, F., Béland, R., Lassonde, M., & Gallagher, A. (2016). Distinct hemispheric specializations for native and non-native languages in one-day-old newborns identified by fNIRS. *Neuropsychologia*, *84*, 63-69. https://doi.org/10.1016/j.neuropsychologia.2016.01.038

Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics*, *60*(6), 926–940. https://doi.org/10.3758/BF03211929

Vergara-Martínez, M., Perea, M., Marín, A., & Carreiras, M. (2011). The processing of consonants and vowels during letter identity and letter position assignment in visual-word recognition: An ERP study. *Brain and Language*, *118*(3), 105–117. https://doi.org/10.1016/j.bandl.2010.09.006

Vihman, M., & Majorano, M. (2017). The role of geminates in infants' early word production and word-form recognition. *Journal of Child Language*, *44*(1), 158–184. https://doi.org/10.1017/S0305000915000793

Vihman, M. M., Nakai, S., DePaolis, R. A., & Hallé, P. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, *50*(3), 336-353. https://doi.org/10.1016/j.jml.2003.11.004

Vihman, M. M., Thierry, G., Lum, J., Keren-Portnoy, T., & Martin, P. (2007). Onset of word form recognition in English, Welsh, and English-Welsh bilingual infants. *Applied Psycholinguistics*, *28*(3), 475–493. https://doi.org/10.1017/S0142716407070269

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*(4), 325–329. https://doi.org/10.1111/1467-9280.00064

Von Holzen, K., & Nazzi, T. (2020). Emergence of a consonant bias during the first year of life: New evidence from own-name recognition. *Infancy*, *25*(3), 319-346. https://doi.org/10.1111/infa.12331

Von Holzen, K., Nishibayashi, L. L., & Nazzi, T. (2018). Consonant and vowel processing in word form segmentation: An infant ERP study. *Brain Sciences*, *8*(2), 24. https://doi.org/10.3390/brainsci8020024

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates preference for speech. *Child Development*, *81*(2), 517–527. https://doi.org/10.1111/j.1467-8624.2009.01412.x

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x

Wanrooij, K., Boersma, P., & Van Zuijen, T. L. (2014). Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study. *Frontiers in Psychology*, *5*, 77. https://doi.org/10.3389/fpsyg.2014.00077

Weatherhead, D., & White, K. S. (2017). Read my lips: Visual speech influences word processing in infants. *Cognition*, *160*, 103-109. https://doi.org/10.1016/j.cognition.2017.01.002

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Werker, J. F., Byers-Heinlein, K., & Fennell, C. T. (2009). Bilingual beginnings to learning words. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3649-3663. https://doi.org/10.1098/rstb.2009.0105

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, *52*(1), 349–355. https://doi.org/10.2307/1129249

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*(5), 672-683. https://doi.org/10.1037/0012-1649.24.5.672

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, *37*(2), 278–286. https://doi.org/10.1037/h0080725

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49-63. https://doi.org/10.1016/S0163-6383(84)80022-3

Werner, L.A., Marean, G.C., Halpin, C.F., Spetner, N.B., & Gillenwater, J.M. (1992). Infant auditory temporal acuity: gap detection. *Child Development*, *63*, 260–272. https://doi.org/10.1111/j.1467-8624.1992.tb01625.x

Wewalaarachchi, T. D., & Singh, L. (2020). Vowel, consonant, and tone variation exert asymmetrical effects on spoken word recognition: Evidence from 6-year-old monolingual and bilingual learners of Mandarin. *Journal of Experimental Child Psychology*, *189*, 104698. https://doi.org/10.1016/j.jecp.2019.104698

Wewalaarachchi, T. D., Wong, L. H., & Singh, L. (2017). Vowels, consonants, and lexical tones: Sensitivity to phonological variation in monolingual Mandarin and bilingual English–Mandarin toddlers. *Journal of Experimental Child Psychology*, *159*, 16-33. https://doi.org/10.1016/j.jecp.2017.01.009

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665-679. https://doi.org/10.1016/j.jml.2011.12.010

Wiener, S. (2020). Second language learners develop non-native lexical processing biases. *Bilingualism: Language and Cognition*, *23*(1), 119–130. https://doi.org/10.1017/S1366728918001165

Wood, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *The Journal of the Acoustical Society of America*, *60*(6), 1381-1389. https://doi.org/10.1121/1.381231

Wozniak-Kaelin, V. D., & Jackson, P. L. (1979). Visual vowel and diphthong perception from two horizontal viewing angles. *Journal of Speech & Hearing Research*, *22*(2), 354–365. https://doi.org/10.1044/jshr.2202.354

Yavaş, M. (2011). *Applied English Phonology*. West Sussex, UK: John Wiley & Sons.

Yeung, H. H., Chen, L. M., & Werker, J. F. (2014). Referential labeling can facilitate phonetic learning in infancy. *Child Development*, *85*(3), 1036-1049. https://doi.org/10.1111/cdev.12185

Yeung, H. H., & Nazzi, T. (2014). Object labeling influences infant phonetic learning and generalization. *Cognition*, *132*(2), 151-163. https://doi.org/10.1016/j.cognition.2014.04.001

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*(2), 234-243. https://doi.org/10.1016/j.cognition.2009.08.010

Yi, A., Wong, W., & Eizenman, M. (2013). Gaze patterns and audiovisual speech enhancement. *Journal of Speech, Language, and Hearing Research*, *56*(2), 471–480. https://doi.org/10.1044/1092-4388(2012/10-0288)

Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press.

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, *15*(4), 420–433. https://doi.org/10.1111/j.1532-7078.2009.00024.x

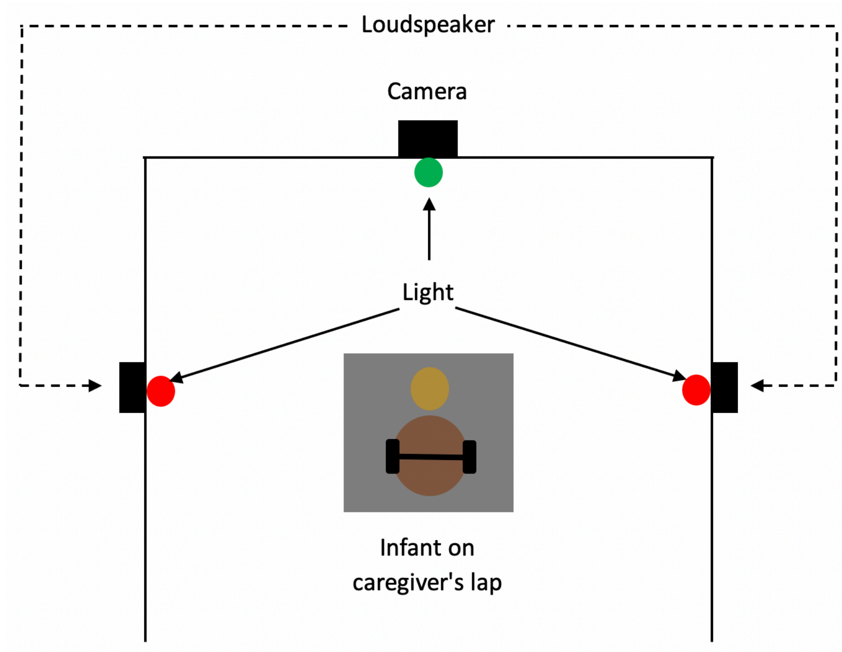
Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, *12*(5), 798–814. https://doi.org/10.1111/j.1467-7687.2009.00833.x

Yuan, Y., Wayland, R., & Oh, Y. (2020). Visual analog of the acoustic amplitude envelope benefits speech perception in noise. *The Journal of the Acoustical Society of America*, *147*(3), EL246-EL251. https://doi.org/10.1121/10.0000737

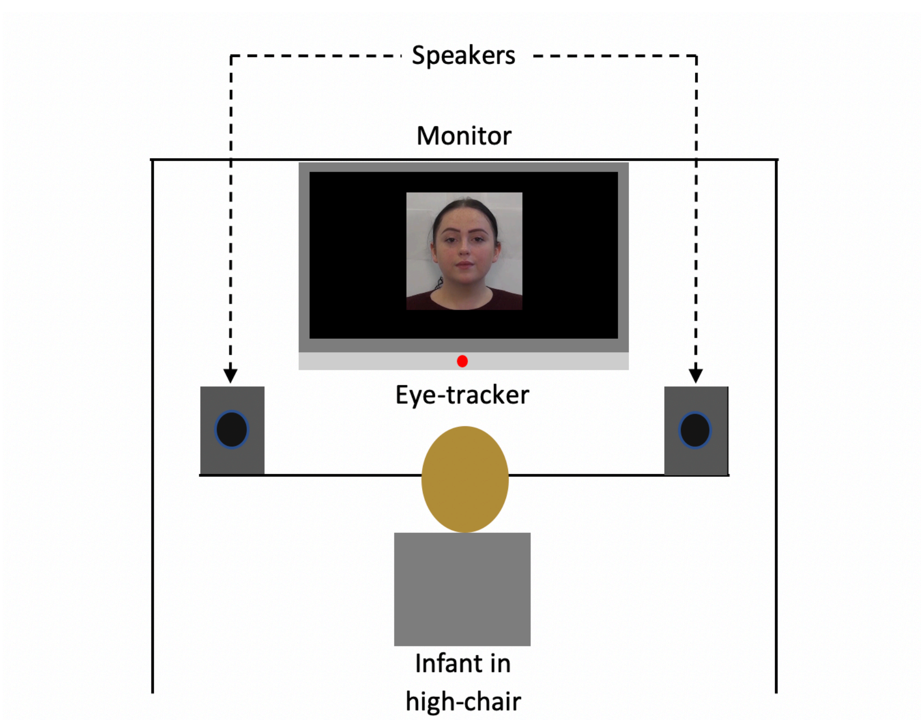
Zesiger, P., & Jöhr, J. (2011). Phonological representations of words in young children. *Childhood*, *3*, 293-309.

8 Appendices

Appendix A. Experimental setup used in Experiments 1 to 3.



Appendix B. Experimental setup used in Experiments 4 and 5.



Appendix C. Pearson correlation (*r*), *p* values, and Bayes Factor (BF) between infants' listening preferences and lexical factors in Experiment 4.

	Correct vs. Consonant			Correct vs. Vowel			Consonant vs. Vowel			PTLT		
Lexical measure	*r*	*p*	BF	r	*p*	BF	*r*	*p*	BF	*r*	*p*	BF
CDI	-.12	.56	.29	-.07	.75	.26	-.22	.29	.42	-.23	.26	.45
Word list	.24	.26	.46	-.02	.94	.25	-.28	.17	.60	-.08	.69	.27
Consonant tiers	-.15	.46	.32	-.06	.79	.26	-.24	.24	.48	-.21	.32	.40
Vowel tiers	-.13	.55	.29	-.06	.77	.26	.22	.30	.41	-.25	.23	.49

Appendix D. Pearson correlation (*r*), *p* values, and Bayes Factor (BF) between infants' listening preferences and lexical factors in Experiment 5.

	Correct vs. Consonant			Correct vs. Vowel			Consonant vs. Vowel			PTLT		
Lexical measure	*r*	*p*	BF	*r*	*p*	BF	*r*	*p*	BF	*r*	*p*	BF
CDI	-.26	.29	.50	-.04	.87	.26	-.05	.80	.26	.01	.96	.25
Consonant tiers	.09	.67	.28	.04	.87	.30	.12	.56	.30	.29	.18	.60
Vowel tiers	.19	.38	.37	-.02	.91	.29	.12	.57	.29	.002	.99	.25