SENSORIMOTOR INFLUENCES ON SPEECH PERCEPTION IN INFANCY

by

ALISON JEANNE GREUEL

B.A., University of Wisconsin-Madison, 2008 M.A., The University of British Columbia, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2014

© Alison Jeanne Greuel, 2014

Abstract

The multisensory nature of speech, and in particular, the modulatory influence of one's own articulators during speech processing, is well established in adults. However, the origins of the sensorimotor influence on auditory speech perception are largely unknown, and require the examination of a population in which a link between speech perception and speech production is not well-defined; by studying *preverbal* infant speech perception, such early links can be characterized. Across three experimental chapters, I provide evidence that articulatory information selectively affects the perception of speech sounds in preverbal infants, using both neuroimaging and behavioral measures. In Chapter 2, I use a looking time procedure to show that in 6-month-old infants, articulatory information can impede the perception of a consonant contrast when the related articulator is selectively impaired. In Chapter 3, I use the highamplitude suck (HAS) procedure to show that neonates are able to discriminate and exhibit memory for the vowels /u/ and /i/; however, the information from the infants' articulators (a rounded lip shape) seems to only marginally affect behavior during the learning of these vowel sounds. In Chapter 4, I co-register HAS with a neuroimaging technique – Near Infrared Spectroscopy (NIRS) – and identify underlying neural networks in newborn infants that are sensitive to the sensorimotor-auditory match, in that the vowel which matches the lip shape (/u/)is processed differently than the vowel that is not related to the lip shape (/i/). Together, the experiments reported in this dissertation suggest that even before infants gain control over their articulators and speak their first words, their sensorimotor systems are interacting with their perceptual systems as they process auditory speech information.

Preface

This dissertation is my own original, unpublished work, and I was the senior author on all collaborative projects (Chapters 2-4).

Chapter 1: Introduction

I am the primary author of this chapter, with intellectual contributions and comments from Janet F. Werker, PhD (supervisor).

Chapter 2: Experiments 1 – 4

I am the primary author of this chapter. The research question and study designs were decided in collaboration with D. Kyle Danielson, MSc, Padmapriya Kandhadai, PhD, and Janet F. Werker, PhD (supervisor). I determined and implemented the final experimental designs. I collected all data, conducted all analyses, and wrote the current report. This research is covered under UBC Ethics Certificate B95-0023/H95-80023 (UBC Behavioural Research Ethics Board).

Chapters 3 & 4: Experiments 5 - 7

I am the primary author of these chapters. The research questions and study designs were decided in collaboration with H. Henny Yeung, PhD, and Janet F. Werker, PhD (supervisor). I collected all data, conducted all analyses, and wrote the current reports. This research is covered under UBC Ethics Certificate B02-0575/H02-80575 (Children's and Women's Research Ethics Board).

Chapter 5: General Discussion

I am the primary author of this chapter, with intellectual contributions and comments from Janet F. Werker, PhD (supervisor).

Abstract	ii
Preface	iii
Table of Contents	iv
List of Figures	vii
Acknowledgements	viii
Dedication	ix
Chapter 1 : Introduction	
1.1 Embodied approaches & motor theories of speech perception	
1.2 Sensorimotor influences on speech perception: Evidence from adults	
1.2.1 Neurological evidence for multisensory speech perception in adults	
1.2.2 Some theories related to the onset of sensorimotor influences on speech perception	
1.3 Speech perception in preverbal infants	
1 3 1 Neonatal period	12
1 3 2 Preverbal period	14
1.4 Multisensory speech perception in preverbal infants	
1 4 1 Audiovisual speech processing in infancy	15
1 4 2 Sensorimotor influences on speech perception in infancy	17
1 4 3 Speech production – perception interface in infancy	18
1.5 Building a framework for development of sensorimotor influences on speech perception	on: The
dorsal stream & multisensory integration in speech processing	21
1.5.1 Dorsal stream and correlated articulatory-auditory information	23
1.5.2 Dorsal stream and language processing in infancy	24
1 6 Current experiments	27
Chapter 2 : Sensorimotor influences on speech sound discrimination in 6-month-old	infants 30
2.1 Introduction	
2.1.1 Multisensory speech perception in infancy	30
2.1.2 Sensorimotor influences on speech perception in infancy	
2.1.3 Current experiments: Sensorimotor influences on non-native consonant perception in	infancy
	35
2.2 Experiment 1: Non-native speech perception in 6-month-old English-learning infants	
2.2.1 Method	
2.2.1.1 Participants	37
2.2.1.2 Stimuli	
2.2.1.3 Apparatus	
2.2.1.4 Procedure	
2.2.1.5 Hypotneses	40 41
2.2.2 Results	
2.2.5 Discussion	
2.5 Experiment 2: Ultrasound images of infant tongue contours	
2.3.1 Mielinod	
2.3.1.1 Participants	
2.3.1.2 Apparatus	43 41
2.3.1.4 Data analysis & hypotheses	
$J_{\mathbf{r}} = J_{\mathbf{r}}$	

Table of Contents

2.3.2 Results	
2.3.3 Discussion	
2.4 Experiment 3: Non-native speech perception by English-learning infants with a	n
articulatory-motor perturbation	•••••
2.4.1 Method	
2.4.1.1 Patticipalits	
2.4.1.2 Stimuli & apparatus	
2.4.1.4 Hypotheses	
2.4.2 Results	
2.4.3 Discussion	
2.5 Experiment 4: Non-native speech perception in English-learning infants with a	non-spe
articulatory-motor perturbation	
2.5.1 Method	
2.5.1.1 Participants	
2.5.1.2 Stimuli & apparatus	
2.5.1.3 Procedure	
2.5.1.4 Hypotheses	
2.5.2 Results	
2.5.2.1 Results: Comparison of Experiment 3 and Experiment 4	•••••
2.5.3 Discussion	•••••
2.6 1 Will this shares parent development and in concert with other concert systems	
2.6.1 will this change across development, and in concert with other sensory systems	<i>S · · · · · · · · · · · · · · · · · · ·</i>
2.6.2 Domain-general of domain-specific. Is this special to speech?	
2.0.5 Conclusions	•••••
hapter 3 : Sensorimotor influences on memory for speech sounds in newborr	18
hapter 3 : Sensorimotor influences on memory for speech sounds in newbori 3.1 Introduction	18
hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction	18
hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction 3.1.1 Neonatal speech perception 3.1.2 High Amplitude Suck procedure	18
 Appendix Second control of the seco	18
 Appendix Second Secon	18 Is in neoi
 hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction	15 Is in neoi
 hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction	18 Is in neoi
 hapter 3 : Sensorimotor influences on memory for speech sounds in newborn 3.1 Introduction 3.1.1 Neonatal speech perception 3.1.2 High Amplitude Suck procedure 3.1.3 Current experiment: Testing learning and memory for vowels using HAS 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1.1 Participants 3.2.1.2 Stimuli 	18 Is in neoi
 hapter 3 : Sensorimotor influences on memory for speech sounds in newborn 3.1 Introduction	15 Is in neoi
 hapter 3 : Sensorimotor influences on memory for speech sounds in newborn 3.1 Introduction 3.1.1 Neonatal speech perception. 3.1.2 High Amplitude Suck procedure. 3.1.3 Current experiment: Testing learning and memory for vowels using HAS. 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1.1 Participants. 3.2.1.2 Stimuli 3.2.1.3 Apparatus. 3.2.1.4 Procedure 	15 s in neo
hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction	15 Is in neol
 hapter 3 : Sensorimotor influences on memory for speech sounds in newborn 3.1 Introduction 3.1.1 Neonatal speech perception 3.1.2 High Amplitude Suck procedure 3.1.3 Current experiment: Testing learning and memory for vowels using HAS 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1.2 Stimuli 3.2.1.3 Apparatus 3.2.1.4 Procedure 3.2.2 Results 3.2.1 High amplitude sucks per minute 	15 Is in neol
 hapter 3 : Sensorimotor influences on memory for speech sounds in newborn 3.1 Introduction 3.1.1 Neonatal speech perception 3.1.2 High Amplitude Suck procedure 3.1.3 Current experiment: Testing learning and memory for vowels using HAS 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1.1 Participants 3.2.1.2 Stimuli 3.2.1.3 Apparatus 3.2.1.4 Procedure 3.2.2 Results 3.2.2 I High amplitude sucks per minute 3.2.2 Average suck amplitude 	15 s in neo
 Appler 3 : Sensorimotor influences on memory for speech sounds in newborn 3.1 Introduction 3.1.1 Neonatal speech perception 3.1.2 High Amplitude Suck procedure 3.1.3 Current experiment: Testing learning and memory for vowels using HAS 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1.1 Participants 3.2.1.2 Stimuli 3.2.1.3 Apparatus 3.2.1.4 Procedure 3.2.2 Results 3.2.2 Results 3.2.2 Average suck amplitude 	15 s in neo
 hapter 3 : Sensorimotor influences on memory for speech sounds in newborn 3.1 Introduction 3.1.1 Neonatal speech perception 3.1.2 High Amplitude Suck procedure 3.1.3 Current experiment: Testing learning and memory for vowels using HAS 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1.1 Participants 3.2.1.2 Stimuli 3.2.1.3 Apparatus 3.2.1.4 Procedure 3.2.2 Results 3.2.2 Results 3.2.2 Average suck amplitude 3.3 General discussion 3.3.1 Memory for and discrimination of vowels 	15 s in neo
 hapter 3 : Sensorimotor influences on memory for speech sounds in newborn 3.1 Introduction 3.1.1 Neonatal speech perception. 3.1.2 High Amplitude Suck procedure. 3.1.3 Current experiment: Testing learning and memory for vowels using HAS. 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1 Participants. 3.2.1.2 Stimuli 3.2.1.4 Procedure. 3.2.2 Results 3.2.2 Results 3.2.2 Average suck amplitude 3.3 General discussion 3.3.1 Memory for and discrimination of vowels 3.3.2 Suck amplitude as a dependent measure 	15 s in neo
 hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction 3.1.1 Neonatal speech perception. 3.1.2 High Amplitude Suck procedure. 3.1.3 Current experiment: Testing learning and memory for vowels using HAS. 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1.1 Participants. 3.2.1.2 Stimuli 3.2.1.3 Apparatus. 3.2.1.4 Procedure. 3.2 Results 3.2.2.1 High amplitude sucks per minute. 3.2.2.2 Average suck amplitude. 3.3 General discussion 3.3.1 Memory for and discrimination of vowels. 3.3.2 Suck amplitude as a dependent measure. 3.3 Sensorimotor effects and 'contingency awareness' 	15 s in neo
 hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction 3.1.1 Neonatal speech perception 3.1.2 High Amplitude Suck procedure 3.1.3 Current experiment: Testing learning and memory for vowels using HAS 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1.1 Participants 3.2.1.2 Stimuli 3.2.1.3 Apparatus 3.2.1.4 Procedure 3.2.2 Results 3.2.2.1 High amplitude sucks per minute 3.2.2.2 Average suck amplitude 3.3 General discussion 3.3.1 Memory for and discrimination of vowels 3.3.2 Suck amplitude as a dependent measure 3.3.4 Conclusions 	15 s in neo
 2.6.3 Conclusions hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction 3.1.1 Neonatal speech perception. 3.1.2 High Amplitude Suck procedure. 3.1.3 Current experiment: Testing learning and memory for vowels using HAS. 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1.1 Participants 3.2.1.2 Stimuli 3.2.1.3 Apparatus. 3.2.1.4 Procedure. 3.2.2 Results 3.2.2.1 High amplitude sucks per minute. 3.2.2.1 High amplitude sucks per minute. 3.2.2 Results 3.3.2 Suck amplitude as a dependent measure. 3.3 Sensorimotor effects and 'contingency awareness' 3.3.4 Conclusions 	15 s in neo
 2.6.3 Conclusions hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction 3.1.1 Neonatal speech perception. 3.1.2 High Amplitude Suck procedure. 3.1.3 Current experiment: Testing learning and memory for vowels using HAS. 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1.1 Participants. 3.2.1.2 Stimuli 3.2.1.3 Apparatus. 3.2.1.4 Procedure. 3.2.2 Results. 3.2.2.1 High amplitude sucks per minute. 3.2.2.1 High amplitude sucks per minute. 3.2.2.1 High amplitude sucks per minute. 3.2.2.3 Average suck amplitude. 3.3 General discussion 3.3.1 Memory for and discrimination of vowels. 3.3.2 Suck amplitude as a dependent measure 3.3.3 Sensorimotor effects and 'contingency awareness' 3.3.4 Conclusions 	15 s in neo
 2.6.5 Conclusions hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction 3.1.1 Neonatal speech perception 3.1.2 High Amplitude Suck procedure 3.1.3 Current experiment: Testing learning and memory for vowels using HAS 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1 Participants 3.2.1.2 Stimuli 3.2.1.3 Apparatus 3.2.1.4 Procedure 3.2.2 Results 3.2.2.1 High amplitude sucks per minute 3.2.2 Average suck amplitude 3.3 General discussion 3.3 Sensorimotor effects and 'contingency awareness' 3.3.4 Conclusions 	15 s in neo
 2.0.3 Conclusions hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction 3.1.1 Neonatal speech perception 3.1.2 High Amplitude Suck procedure 3.1.3 Current experiment: Testing learning and memory for vowels using HAS 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1.1 Participants 3.2.1.2 Stimuli 3.2.1.3 Apparatus 3.2.1.4 Procedure 3.2.2.1 High amplitude sucks per minute 3.2.2.1 High amplitude sucks per minute 3.2.2.2 Average suck amplitude 3.3 General discussion 3.3 I Memory for and discrimination of vowels 3.3 Sensorimotor effects and 'contingency awareness' 3.3 4 Conclusions 	15 s in neo
 2.0.3 Conclusions hapter 3 : Sensorimotor influences on memory for speech sounds in newborr 3.1 Introduction 3.1.1 Neonatal speech perception. 3.1.2 High Amplitude Suck procedure. 3.1.3 Current experiment: Testing learning and memory for vowels using HAS. 3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sound 3.2.1 Method 3.2.1 Participants 3.2.1.2 Stimuli 3.2.1.3 Apparatus. 3.2.1.4 Procedure 3.2.2 Results 3.2.2.1 High amplitude sucks per minute. 3.2.2 Average suck amplitude 3.3 General discussion 3.3.1 Memory for and discrimination of vowels 3.3.2 Suck amplitude as a dependent measure 3.3.3 Sensorimotor effects and 'contingency awareness' 3.3.4 Conclusions hapter 4 : Neural networks involved in processing vowel sounds in neonates. 4.1 Introduction	15 s in neo

4.2. Experiment 6: Neural networks involved during speech sound presentation <i>contingent</i> suck behavior	t on 97
4.2.1 Method	
4.2.1.1 Participants	
4.2.1.2 Stimuli & apparatus	
4.2.1.3 Procedure	
4.2.1.4 Data analysis and preparation of matrices	
4.2.2 Results	104
4.2.2.1 Constrained Principal Component Analysis: Familiarization phase	104
4.2.2.2 Constrained Principal Component Analysis: Test phase	108
4.2.2.3 Behavioral data: High amplitude sucks	114
4.2.3 Discussion	115
4.3. Experiment 7: Neural networks involved during speech sound presentation not conting	gent on
	110
4.3.1 Method	115
4.5.1.1 Participants	110
4.5.1.2 Sumul & apparatus	115 110
4.5.1.5 Flotedule	11: 12(
4.5.1.4 Data analysis and preparation of matrices	120
4.3.2 ACSUITS	120
4.3.2.7 Constrained Thicipal Component Analysis. Tammanzation phase	120
4 3 3 Discussion	12
4.5.5 Discussion	12:
4.4.1 Sensorimotor effect: specific to $/u/2$	120
4.4.2 CPCA to analyze event related date in NIPS studies	120
4.4.2 CPCA to analyze event-related data in NIKS studies	125
4.4.5 Conclusions	132
Chapter 5 : General discussion	133
5.1 Summary of experimental chapters	133
5.2 Implications of empirical findings and future directions	139
5.2.1 Sensorimotor influences on (speech) perception—how?	14(
5.2.2 Linking multiple modalities across development: audition vision and motor systems	144
5.2.2 Employed and the systems of a second s	1 1 144
5.3 Conclusions	147
References	148
Appendices	176
Annendix A: Experiment 7 test phase—CPCA	174
Tippenuia II Daperintent / test phase CI CA and	170

List of Figures

Figure 1.1. Brain areas associated with multisensory integration.	
Figure 1.2. Ventral & dorsal language pathways	
Figure 2.1. Images of Hindi speaker producing the phonemes a) $/d/$ and b) $/d/$	
Figure 2.2. Images of the a) flat teether and b) gummy teether	
Figure 2.3. Ultrasound image from live recording (flat teether in the mouth).	
Figure 2.4. Tongue contours of 3 infants comparing gummy teether to no teether	
Figure 2.5. Tongue contours of 3 infants comparing flat teether to no teether	
Figure 2.6. Tongue contours of 3 infants comparing flat teether to gummy teether	49
Figure 2.7. Experiment 3 looking time averages during test trials.	
Figure 2.8. Experiment 4 looking time averages during test trials.	59
Figure 2.9. Average looking times during test trials for Experiments 3 and 4	
Figure 3.1. Familiarization phase HAS data split by familiarization-vowel.	
Figure 3.2. Test phase HAS data split by experimental condition	
Figure 4.1. Ventral and dorsal language pathways in adults and neonates	
Figure 4.2. Representation of NIRS probe placement on a neonate skull.	
Figure 4.3. Experiment 6 familiarization phase—Component 1	106
Figure 4.4. Experiment 6 familiarization phase—Component 2.	108
Figure 4.5. Experiment 6 test phase—Component 1	111
Figure 4.6. Experiment 6 test phase—Component 2.	113
Figure 4.7. Experiment 7 familiarization phase—Component 1	122
Figure 4.8. Experiment 7 familiarization phase—Component 2	123

Acknowledgements

I begin with a heartfelt thank you to my supervisor and mentor, Janet Werker. This document not only represents a personal feat, but it is also a testament to Janet's unabating patience and dedication to my development as a student, researcher, and person. Every step of the way, even as she pushed and challenged me to think more deeply and more critically, Janet reminded me to keep my own goals and happiness in perspective; this is more than I could have asked for in a supervisor. Thank you, Janet. I wouldn't have done this without you.

I thank my committee members, Jim Enns and Rebecca Todd, for their insight and guidance as the ideas in this document came to fruition. I would also like to acknowledge Eric Vatikiotis-Bateson, Bryan Gick, Noriko Yamane, Behnam Molavi, and Todd Woodward, without whom I wouldn't have been able to use such neat equipment, or conduct such challenging analyses, as I did in this dissertation.

Throughout my degree, I have had the pleasure of working with an amazing group of people. I want thank my friends and colleagues in my lab family—Lily May, Alexis Black, Savannah Nijeboer, Samantha Bangayan, Zaineb Waheed, and Nurit Gazit-Gurel—for their honest advice, patience, and support, especially in these last few months. I am utterly grateful for my collaborators—Priya Kandhadai, Kyle Danielson, and Henny Yeung—for reminding me that research is more fun, and more intellectually challenging, when you get to work with such clever, creative people. I'd like to also acknowledge some of my academic and personal role models—Laurie Fais, Afra Foroud, Krista Byers-Heinlein, and Judit Gervain—for providing me with mentorship and exemplifying success. I'm thankful for the support and friendship I found in my stellar coworkers, even if we only worked together for just short periods of time—Neda Razaz-Rahmati, Julia Leibowich, and Emily Chevrier. And finally, I am indebted to the volunteers and research assistants—especially Sharon To, Anthea Pun, and Maria Ho—who spent countless hours contacting the parents and families who graciously participated in the research that follows. All of you helped make this dissertation happen, and for that I feel very fortunate.

I have a wonderful support group outside of the lab who also deserve a huge thank you. My dear friends—especially Annie and Lizzy—thank you for keeping me sane, helping me enjoy the beauty of this city, and providing me with countless memories that still have me smiling. I thank my family for their unwavering love, endless encouragement, and special sense of humor, even after I moved thousands of miles from home—my parents, Jeff and Maureen; my sisters, Ashley and Abby; my grandmother, June; and of course, Grace and Peter, my dearest niece and nephew. I miss you guys everyday.

Finally, I'd like to acknowledge my partner in crime and very best friend, Tom. Somehow you've kept me laughing, and grounded, and honest the whole way, and I love you so much for your support. You've helped me in more ways than you'll ever know, and for that, I thank you.

Dedication

For Peter

"I believe in kindness. Also in mischief. Also in singing, especially when singing is not necessarily prescribed." -- Mary Oliver

Chapter 1 : Introduction

"What's the first thing a baby does when she wants to learn about something? She puts it in her mouth. Why do we think the way she learns language is any different?" a mentor once asked me. The information available to and used by developing infants as they learn about the world around them is of particular interest to developmental and cognitive psychologists alike, and the way infants process and learn about language can ultimately inform theories of perception more generally. In this dissertation, I ask whether preverbal infants, from birth to 6 months of age, recruit information from their sensorimotor systems while processing speech. I provide evidence that, in both behavioral and neuroimaging procedures, sensorimotor information from an infant's articulators does impact speech processing; I discuss the implications of these findings, and place them into the existing research landscape of theories of speech perception, particularly those that posit a linkage between the motor systems and speech perception systems.

1.1 Embodied approaches & motor theories of speech perception

The way that I perceive a scene, an action, or an event usually involves information from several domains. At any given moment, my sensory systems are taking in information that I see, hear, feel, and even taste and smell, resulting in a unified percept of the event. The same is true for the way I process speech, a signal that has largely been assumed to be auditory in nature; it is now well documented that humans recruit information from outside the auditory system when listening to and processing speech (Campbell & Dodd, 1980; Massaro, 1998, 2004; Navarra, Yeung, Werker, & Soto-Faraco, 2013). Research concerning the multisensory nature of speech perception has predominately focused on the role of visual speech information, and how it

affects the auditory speech we perceive (Bishop & Miller, 2009; Kuhl, Williams, & Meltzoff, 1991; Tuomainen, Anderson, Tiippana, & Sams, 2005). As we experience during a conversation in a noisy environment, watching our conversational partner's face can help disambiguate the person's speech, increasing the intelligibility of the speech signal (Sumby & Pollack, 1954; Summerfield, 1979). When auditory and visual speech information conflict, as demonstrated by the McGurk effect, the perceptual system mandatorily combines the discrepant auditory and visual signals into a unified speech percept (Kislyuk, Mottonen, & Sams, 2008; McGurk & MacDonald, 1976). For example, when a subject listens to the syllable /ba/ and simultaneously sees a speaker mouth the syllable /ga/, the resultant unified percept is the syllable /da/. Visual speech can also enhance the perception of a non-native phonemic contrast, one that is typically difficult to discriminate in auditory-only tasks (Navarra & Soto-Faraco, 2007). Together, these results suggest that visual speech information plays an important, and even deterministic, role in the speech percept.

However, humans have more than auditory and visual information available to them as they perceive speech: they have articulatory-motor (proprioceptive)¹ access, or the ability to simultaneously produce and perceive speech. Motor theories of perception and theories of embodied cognition purport that we recruit information from the motor system during perception generally (Barsalou, 2008; Wilson, 2002), and during the perception of speech specifically (Fowler, Galuntucci, & Saltzman, 2003). As argued by the proponents of motor theories of speech perception, particularly in the 'strong' version of the theory, a link between the speech perception and speech production systems may be such that humans perceive speech as a series of articulatory gestures, rather than as a series of acoustic signals, and that perception occurs in a

¹ Throughout this document, I use 'articulatory-motor' and 'sensorimotor' interchangeably.

specialized module that is separate from the perception of other, non-linguistic auditory stimuli (Liberman & Mattingly, 1985). An alternative view, one that claims speech perception is strictly an auditory process and is based in no way on the motor system (Ohala, 1996), draws on evidence that a) sounds of the world's languages differ most in their acoustic properties, rather than how they are articulated by native speakers (see Ohala, 1996 for a review of evidence from linguistics); and b) preverbal infants and non-human animals (including chinchillas and Japanese quail) can differentiate speech sounds which they themselves are not able to produce (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Kluender, Diehl, & Killeen, 1987; Kuhl & Miller, 1975; Werker & Tees, 1984). This debate is mentioned here only to highlight the range of opinions that exist regarding a role for the articulatory system in the perception of speech.

A more nuanced view of the sensorimotor influence on speech perception is one that argues for a modulatory effect on the perception of speech—this view states that while the sensorimotor system (and information from the articulators) is not required for the perception of speech information, it does modulate speech perception (Hickok, Houde, & Rong, 2011, described below). In the following section, I review some of the recent behavioral findings that showcase these modulatory effects.

1.2 Sensorimotor influences on speech perception: Evidence from adults

Evidence in support of a link between the auditory and sensorimotor systems during speech perception is mainly found in adult literature. For instance, when an adult silently articulates a syllable (such as /ka/) while simultaneously listening to a different syllable (such as /pa/), the discordant production-perception information decreases the ability to identify the auditorily presented syllable (Sams, Möttönen, & Sihvonen, 2005). In comparing the disruption of articulatory information versus visual speech information during speech perception, the

articulatory-induced interference during speech perception tasks is articulator-specific, while the visually-induced interference is not (Mochida, Kimura, Hiroya, Kitagawa, Gomi, & Kondo, 2013). Further, imagining the production of a speech sound can affect how one perceives the external, actual speech sound (Scott, Yeung, Gick, & Werker, 2013). Even when some of the articulators are mechanically stretched in a way that resembles the deformation of skin during the natural production of speech (in this case, in the shape and opening of the lips), perception of speech is altered (Ito, Tiede, & Ostry, 2009).

Speech information can also be experienced haptically on the skin, and the tactile or haptic information can affect the way adults perceive speech. For example, when adults feel another person mouth a syllable on their skin while they listen to a different, synchronously-presented auditory token, the haptic speech information affects perception of the sound (Fowler & Dekle, 1991). In addition, adults can integrate naturalistic tactile information in the form of a small puff of air during the perception of auditory speech (Gick & Derrick, 2009). During natural speech production, some sounds produce small bursts of air (these sounds are referred to as 'aspirated', and include the /p/ in the word 'pot'); when tiny air bursts are applied to the skin while listening to speech, subjects are more likely to hear the sound as an aspirated one, even if it was un-aspirated (such as the /b/ in the word 'bog'). Thus, subjects may incorrectly report hearing /pa/, even if the sound was /ba/. Together, these results suggest that haptic and tactile information, like visual and sensorimotor information, can influence the perception of auditory speech.

1.2.1 Neurological evidence for multisensory speech perception in adults

Studying the underlying neural activity during perceptual tasks allows researchers to ask questions about integration and processing that would otherwise be unanswerable using

behavioral techniques. In neurophysiological literature, *multisensory integration* is defined as a process that combines information from cross-modal stimuli. At the level of a single neuron, multisensory integration has occurred when the cross-modal stimuli (i.e., auditory and visual information) evoke responses that are *different* from the responses evoked by the individual, component stimuli (auditory or visual information) (Stein & Stanford, 2008). This could result in either a supra-additive response, in which the combined multisensory information leads to a greater response than the sum of the two component responses, or a subadditive response, in which the multisensory response is less than the sum of the two component responses. A midbrain structure known as the Superior Colliculus (SC) receives inputs from visual, auditory, and somatosensory areas, and contain multisensory integrative cells (MSI), the most extensively studied single cells for integrating information cross domains (Calvert, 2001; Meredith & Stein, 1986; Miller & D'Esposito, 2005; Stein & Stanford, 2008). In most cases, spatially congruent multisensory information results in supra-additive responses from MSI cells, while spatially disparate multisensory information leads to sub-additive responses in the MSI cells (Calvert, 2001).

Considering higher level neural areas, during the perception of multisensory information or events, differences in activation have been identified in the Superior Temporal Sulcus (STS), which has been implicated in the processing of biological motion and intelligible speech (Miller & D'Esposito, 2005); in processing multisensory speech information (particularly audiovisual speech), activation in the STS is left-lateralized, as is seen in auditory speech processing tasks (Calvert, Campbell, & Brammer, 2000). When subjects are exposed to semantically congruent audiovisual speech (compared to auditory or visual alone), the left STS exhibits supra-additive activation, while semantically incongruent (mismatched) audiovisual speech results in a sub-

additive response in the left STS (Calvert et al., 2000). In addition, the response enhancement in audiovisual speech is also accompanied by a decrease in unimodal activity in the auditory cortex, compared to auditory-only speech, as seen in research using ERP (Besle, Fort, Delpuech, & Giard, 2004). Therefore, it seems that the response enhancement or depression found in MSI cells at the level of individual neurons also applies to higher level neural areas, and may be a general characteristic of multisensory integration (see also Wright, Pelphrey, Allison, McKeown, & McCarthy, 2003) (see Figure 1.1 for pictorial representation of cortical areas, from Friederici, 2011).



Figure 1.1. Brain areas associated with multisensory integration. From Friederici, 2011.

In recent years, large advances have occurred in our understanding of the neural underpinnings and processing of sensorimotor speech information (Price, 2010); studies continually provide evidence that premotor and motor areas are involved in speech perception (Iacoboni, 2008), and even have a specific, causal role in speech comprehension that is separate from speech production (D'Ausilio, Pulvermuller, Salmas, Bufalari, Begliomini, & Fadiga, 2009). The ventral premotor cortex is activated bilaterally when subjects listen to monosyllables, but not when they produce the same sounds (Wilson, Saygin, Sereno, & Iacoboni, 2004). Further, when the left motor cortex (specifically in the area implicated in representation of the lips) is disrupted by repetitive transcranial magnetic stimulation (rTMS), the ability to categorically perceive sounds that involve lip closure (such as /ba/ or /pa/) is impaired, but not the perception of sounds that do not require lip closure (such as /da/ or /ga/) (Mottonen & Watkins, 2009). Evidence has also demonstrated articulatory-motor influences during audiovisual speech processing. Auditory and visual speech, even when presented independently, excite the related, articulatory-specific area of the motor cortex during speech perception (Watkins, Strafella, & Paus, 2003). In addition, while experiencing McGurk stimuli of an illusory /ta/ (watching a speaker mouth /ka/ while listening to an audio /pa/), the resultant cortical activation is more similar to activation patterns during an *actual* audiovisual /ta/, compared to the patterns evoked by either audiovisual /pa/ or audiovisual /ka/; these findings demonstrate that audiovisual speech causes a 'motor plan' for the production of the illusory syllable, and results in the illusory percept (Skipper, van Wassenhove, Nusbaum, & Small, 2007). Therefore, not only is the motor system recruited generally in the perception of speech, but it also seems that there is specificity in activation of the motor areas (Pulvermuller, Huss, Kherif, Martin, Hauk, & Shtyrov, 2006).

Not surprisingly, speech production and silent articulation also influence neural areas involved in speech perception. A region in the temporal-parietal junction known as the planum temporale contains a specific area that has been implicated in the integration of sensory (auditory) and motor (production) speech information: area Spt (Hickok, Buchsbaum,

Humphries, & Muftuler, 2003; Hickok, Okada, & Serences, 2009). In this area, the pattern of activation differs during strictly speech perception (listening) tasks, compared to speech production-perception tasks (covert rehearsal while listening), suggesting that this area is specifically implicated in sensorimotor integration of auditory and articulatory information. Further, activation is greater in the superior temporal and inferior parietal cortices when subjects silently articulate sounds while listening to speech compared to when they produce speech (Agnew, McGettigan, Banks, & Scott, 2013).

Separable effects of production and perception have also been shown in tasks that involve non-native speech sounds: while subjects produce words in a non-native language, there is greater activation in the auditory regions and a subsequent increase in activity in articulatory regions (which the authors refer to as auditory feedback) compared to activation during the production of sounds from their native language (Parker Jones, Seghier, Kawabata, Duncan, Leff, Green, & Price, 2013). In contrast, during perception tasks, non-native (second-language) speakers of English show greater activation to a native-English contrast (the /r/-/l/ distinction) in cortical areas involved in articulatory-auditory mapping (including Broca's area, the planum temporale, and area Spt) compared to native English speakers, who show greater activation in auditory areas (Callan, Jones, Callan, & Akahane-Yamada, 2004). In a different perceptual task, while subjects passively listen to non-native, unfamiliar phonemes that vary in articulatory difficulty, a larger signal change can be found in both temporal (auditory) and precentral (motor) areas compared to passively listening to native phonemes (Wilson & Iacoboni, 2006). However, neural activity is related to difficulty of articulation only in the temporal (auditory) areas while listening to non-native phonemes, not the motor areas; the authors suggest that the motor system creates 'top-down' internal models of speech sounds, while the auditory system compares these

internal models (and their possible acoustic consequences) to the acoustic input. Indeed, research has shown that the motor system influences and mediates speech recognition while categorizing speech sounds (Sato, Grabski, Glenberg, Brisebois, Basirat, Menard, & Cattaneo, 2011). Together, these results suggest that the articulatory-auditory mapping cortical areas may play a role in disambiguating difficult speech sound contrasts.

Taking into account the plentiful data showing which cortical areas experience sensorimotor influences during speech perception and speech production, it has been suggested that these influences are modulatory, not required, for perception to occur (Hickok, Houde, & Rong, 2011). In their review, Hickok and colleagues (2011) argue that the neural regions described above, particularly the areas in the planum temporale, play a role in the auditory feedback control of speech production. While these areas are activated during passive listening, the authors suggest that the motor system's influence on the process of speech perception is a top-down, modulatory one; therefore, under some circumstances (including those described above), forward predictions from the motor speech system can influence how listeners perceive other people's speech. They posit that these regions largely support speech production, as "motor acts aim to hit sensory targets" (Hickok, 2012, page 396), and that the neural activation in the temporal-parietal areas (and the modulatory influence of the motor system) found during perception tasks is largely a byproduct of this primary function. However, these researchers do not discuss the establishment of such 'forward predictions'; the modulatory effects of the motor system may become evident in adult speech perception studies because of the extensive experience adults have accrued by both perceiving *and* producing their native language. I now discuss one explanation that has been advanced for the establishment of the speech productionperception linkage: mirror neurons.

1.2.2 Some theories related to the onset of sensorimotor influences on speech perception

When and how do these modulatory effects begin? Some argue that mirror neurons—a type of neuron that fires both during the performance of an action and during the perception of another performing the same action (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996)—explain the sensorimotor link to speech perception, particularly the link suggested by the 'strong' version of the motor theory of speech perception (Rizzolatti & Arbib, 1998; see Cook, Bird, Catmur, Press, & Heyes, 2014 and associated commentary for a review on current findings in mirror neuron literature). Mirror neurons in non-human primates have been found in the area of the cortex that is arguably homologous to Broca's area in humans (Petrides, Cadoret, & Mackey, 2005), and some researchers posit that a similar mirror system in humans could play an important, necessary role in speech perception. However, as Lotto and colleagues (Lotto, Hickok, & Holt, 2009) point out, the motor theory most commonly associated with the mirror neuron argument goes beyond simply proposing a link between perception and production: it claims that speech motor planning is *necessary* for speech perception, and the linguistic system is a completely separate perceptual process (a module). Because of this, and the fact that mirror neurons have been discovered and fire during non-verbal gestures, the mirror neuron theory is at odds with the tenets of the motor theory of speech perception. In addition, the mirror neuron-motor theory argument provides no explanation of how or when the sensorimotor-auditory link in speech perception forms.

Considering, though, the modulatory effects of sensorimotor information on the perception of speech that have been described earlier in this chapter, and the *idea* of neurons that become sensitive to both sensorimotor (articulatory) and auditory speech information at some point in development, a model has been proposed that learns to couple the motor movements during productive behavior (including productions during infant babbling behavior) and their

auditory outputs via Hebbian connections² (Westermann & Miranda, 2004). In the model, when activation in the motor 'map' co-varies with the activation in the auditory 'map', this creates a connection between the motor movement and auditory unit. Over time, the model develops a set of highly correlated responses between the motor movements and sound pairs; thus, Hebbian connections form between highly weighted motor – auditory unit pairs, causing neurons to fire in response to either the production or perception of a speech sound. Thus, this sensorimotor-auditory coupling model assumes that the linkage between motor movements and auditory speech during speech perception tasks is one that depends on experience, and relies heavily on the development of the speech production system. While such a model of sensorimotor-auditory coupling represents just one of the many theories concerning the link between perception and production of speech (see Kroger, Birkholz, & Lowit, 2010; Schwartz, Basirat, Menard, & Sato, 2012; Tourville & Guenther, 2011; Warlaumont, Westermann, Buder, & Oller, 2013), it tells a developmental story that is pertinent to this discussion (Iverson, 2010).

In order to identify whether sensorimotor influences on speech perception are, in fact, due to the developing linkage between production and perception, it is necessary to investigate the sensorimotor nature of speech processing in a population in which language-production experience is limited. The following section features a discussion of speech perception skills in such a population – preverbal infants; I first describe some of the primary milestones of speech perception that occur within the first few months of life, and then turn to and discuss the current literature on multisensory speech processing in infancy.

² I include this model here as an example of a theory that provides a developmental account of the linkage between sensorimotor and auditory speech systems. Whether or not the connections are formed via Hebbian or associative mechanisms is a distinction on which I remain agnostic (see Cooper, Cook, Dickinson, & Heyes, 2013; Cook, Bird, Catmur, Press, & Heyes, 2014, for a discussion).

1.3 Speech perception in preverbal infants

Because infants become native perceivers of their language(s) within the first year of life (Kuhl, 2004; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Gervain, 2013; Werker & Tees, 1984), it is important that the necessary input and properly functioning perceptual systems are available in order for the language systems to develop in full (Saffran, Werker, & Werner, 2006). Some of the accomplishments that follow becoming a native language perceiver include the abilities to learn words, grow a lexicon, and communicate successfully with other members of the language group; perceiving the sounds of the native language is the first step to becoming a successful language user (Tsao, Liu, & Kuhl, 2004; Yeung & Werker, 2005). Language can be processed at many levels—phonemes (speech sounds) and syllables, prosodic information (rhythm and stress), and syntax (grammar); each level has different milestones associated with it across development (Gervain & Mehler, 2010). For the purposes of this dissertation, I focus on the perception of speech sounds, known as phonemes, the perceptual biases newborns have at their disposal, as well as how these capabilities change across the first few months of life; as such, this is not an exhaustive review of infant speech perception literature (see Houston, 2005).

1.3.1 Neonatal period

At birth, neonates have already accrued experience with auditory speech information in utero, as the auditory system is functional by 26 weeks of gestation (Eisenberg, 1976; Graven & Brown, 2008; Moore & Linthicum, 2007). While the speech signals processed by the fetal ear are typically low-frequency and are akin to low-pass filtered sounds, aspects of pitch, rhythm, and some phonetic information can be transmitted from the mother's voice through the uterus (Lecanuet & Granier-Deferre, 1993; Querleu, Renard, Versyp, Paris-Delrue, & Crèpin, 1988).

Indeed, neonates have a preference to listen to their own mother's voice over the voice of an unfamiliar female shortly after birth (DeCasper & Fifer, 1980). Concerning general speech perception capacities at birth, newborn infants not only prefer human speech over non-speech (Vouloumanos & Werker, 2007a), but they also display a familiarity to, preference for, and ability to discriminate their native language(s) from other, rhythmically dissimilar languages (Byers-Heinlein, Burns, & Werker 2010; Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988; Moon, Cooper, & Fifer, 1993). Neuro-imaging studies have identified differential processing for speech over non-speech (Peña, Maki, Kovačić, Dehaene-Lambertz, Koizumi, Bouquet, & Mehler, 2003) as well as for a familiar versus an unfamiliar language (May, Byers-Heinlein, Gervain, & Werker, 2011; Minagawa-Kawai, van der Lely, Ramus, Sato, Mazuka, & Dupoux, 2011).

In addition to these general biases for human speech, newborn infants also have the capability to distinguish particular aspects of speech information. Concerning the perception of phonetic information, newborns exposed to a set of consonant-vowel (CV) syllables (such as /ba/, /bi/, and /bo/) can identify a change in vowel type (addition of /bu/) or an unfamiliar CV syllable (addition of /du/), but not a change in consonant type alone (addition of /da/), suggesting an early capacity to encode small changes in phonetic information (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Kujala, Huotilainen, Hotakainen, Lennes, Parkkonen, Fellman, & Näätänen, 2004). In their seminal work, Eimas and colleagues (Eimas et al., 1971) determined that infants as young as 1 month of age treat speech sounds categorically: when exposed to sounds that varied on a continuum of voiced to voiceless bilabial stops (from /ba/ to /pa/, respectively), infants were more likely to detect a change in sound when it *crossed* the adult phonemic category boundary, rather than when the change occurred *within* the category. This

capability to categorically perceive speech sounds occurs very quickly in processing; as measured with ERP, infants listening to syllables that differ by their initial consonant (/ba/ vs /ga/) can recognize the consonant change in 400 ms (Dehaene-Lambertz & Dehaene, 1994). Taken together, the evidence from infants at birth to one-month of age shows a capacity of the auditory speech perception system to distinguish small phonetic changes, specifically in a categorical manner.

1.3.2 Preverbal period

The ability to categorically perceive speech sounds from early in life has even been shown for speech sounds never before heard (Werker & Lalonde, 1988; Werker & Tees, 1984). As the infant gains more experience with her native language, the speech perception system becomes tuned to her particular linguistic environment, and her initially general phonetic discrimination ability becomes more specific (Kuhl et al., 1992; Werker & Tees, 1984). For example, in English, the contrast between /r/ and /l/ is meaningful—native English speakers can easily tell the difference between the words 'rake' and 'lake'. In Japanese, however, the $\frac{r}{-l}$ distinction is not meaningful, and to a native Japanese speaker, the difference between 'rake' and 'lake' is difficult to perceive (indeed, this difference in perception begins in the first year of life, Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson, 2006). A common speech contrast used to study the development of this perceptual tuning is found in the Hindi language: the voiced stop consonants $\frac{d}{d}$ (a front, dental stop consonant) and $\frac{d}{d}$ (a back, retroflexed stop consonant). As in the 'rake'-'lake' example for English and Japanese speakers, the words /dal/ (which translates to 'lentil') and /dal/ (which translates to 'branch') are easily perceived by native Hindi speakers; to English speakers, the difference between these words is difficult to discriminate as the two stop consonants assimilate to the English /d/, which is produced at the alveolar ridge. Englishlearning infants are able to easily discriminate /d/ and /d/ until 8-10 months of age, at which point the contrast becomes difficult to perceive. This perceptual tuning reflects the fact that English does not treat these two sounds as phonemes, and as such, these sounds are not distinguished in the ambient language by native English speakers (Werker, Gilbert, Humphrey, & Tees, 1981).

By 6 months of age, vowel perception is becoming native-language specific (Kuhl et al., 1992; Polka & Werker, 1994), while consonant perception only starts becoming native-like between 8 and 10 months of age (e.g., Werker & Tees, 1984). The 'perceptual attunement' of speech sound discrimination identifies the importance of early experience with the native language in shaping the speech perception systems in infants; in the next section, I consider the other kinds of sensory information which aid in speech perception tasks, as well as how they impact the developmental trajectory of becoming a native perceiver.

1.4 Multisensory speech perception in preverbal infants

Although the term 'multisensory' implies the integration of information across multiple senses, the majority of the research conducted on infant multisensory speech perception has focused on the auditory and visual domains (see Soto-Faraco, Calabresi, Navarra, Werker, & Lewkowicz, 2012). I begin this section with a review of the audiovisual speech perception capabilities evident in infancy, and follow with a discussion of the current (and limited) literature on the impact of the sensorimotor system on infant speech perception.

1.4.1 Audiovisual speech processing in infancy

Infants are able to integrate auditory and visual speech within the first few months of life, well before they are able to produce speech. Beginning at two months of age, infants correctly match a speaker's face with the corresponding auditory speech information in the speaker's

voice (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999, 2003), and even newborn infants show some sensitivity to the correspondence between visual and auditory speech information (Aldridge, Braga, Walton, & Bower, 1999). Similar to an effect found in adults, 6month-old infants use visual speech information to disambiguate a difficult auditory speech sound contrast (Teinonen, Aslin, Alku, & Csibra, 2008). In addition, there is evidence that preverbal infants integrate disparate auditory and visual speech information into a unified speech percept during a McGurk task (Burnham & Dodd, 2004), and can do so as early as 2 months of age as measured using ERP (Bristow, Dehaene-Lambertz, Mattout, Soares, Gliga, Baillet, & Mangin, 2008; Kushnerenko, Teinonen, Volein, & Csibra, 2008). However, in behavioral studies, integration of disparate auditory and visual speech information can only be identified under certain conditions (Desjardins & Werker, 2004), suggesting it may be related to the attentional demands required to process McGurk stimuli (Alsius, Navara, Campbell, & Soto-Faraco, 2005).

As is the case with auditory speech perception, the audio-visual speech processing system also develops to reflect the infant's linguistic environment: by 11 months of age, infants can no longer match a non-native speaking face with the corresponding non-native auditory speech information, but they maintain the ability to do so for native speakers and corresponding native speech sounds (Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009). Therefore, although infants are readily able to use visual information when perceiving speech even in the first few days after birth, these results suggest that the integration of auditory and visual speech information is also affected by experience with language, and becomes native-language-specific by the end of the first year of life.

1.4.2 Sensorimotor influences on speech perception in infancy

As mentioned previously, the involvement of the sensorimotor system during auditory speech perception remains largely unstudied. At birth, newborns show some evidence of an ability to match the movement of their own articulators with corresponding auditory speech sounds; newborns show more mouth-opening in response to /a/ sounds, and more lip-closure in response to /m/ sounds (Chen, Striano, & Rakoczy, 2004). However, in the Chen et al. (2004) study, visual access to the speaker was not controlled, which means many newborns also watched the speaker produce the sounds. Indeed, audiovisual presentation of the vowels /a/ and /i/ results in more articulatory imitation by neonates than does watching a face articulate in the absence of sound (Coulon, Hemimou, & Streri, 2013). Further, within the first month of life, infants can recognize intermodal matches between objects that were first tactually explored, and later visually presented (Meltzoff & Borton, 1979), implying that neonates do possess intermodal mapping capabilities that involve the visual system (see also Meltzoff & Moore, 1977; Slater, Brown, Hayes, & Quinn, 1999). By 4 months of age, infants vocally imitate vowel sounds in response to audio-visually presented vowels (Kuhl & Meltzoff, 1982; 1996; Patterson & Werker, 1999), suggesting that the early articulatory imitations in response to audiovisual speech may relate to vocal imitations that become evident a few months later.

In the first experimental investigation of the role infants' articulators play during speech perception tasks (by manipulating the shape of the infants' mouths), Yeung and Werker (2013) showed that the shape of 4-month-old infants' lips can influence audio-visual speech processing. Infants experienced an audiovisual speech matching task, where they watched and listened to speakers produce /u/ or /i/; during the study, infants' lip shapes were controlled either by a soother (which rounds the lips) or a flat, wide teether (which spreads the lips). These two lip

configurations correspond to the sounds /u/ and /i/, respectively. While infants correctly looked at the face that matched what they heard when there was nothing in their mouths, their behavior was different when their lip configurations were controlled. During the task, infants in the lip-rounding condition looked more to the /i/ face when they heard the /u/ sound, while infants in the lip-spreading condition looked more to the /u/ face when they heard the /i/ sound. The manipulation of lip configuration resulted in a contrast effect, which suggests that the sensorimotor system can affect audiovisual speech perception in preverbal infants (Yeung & Werker, 2013; see Hamilton, Wolpert & Frith, 2004, for evidence of a contrast effect on perception in adults).

1.4.3 Speech production – perception interface in infancy

However, infants are not only listening to speech in the first few months of life, but they are also beginning to produce a series of primitive speech sounds; the onset of canonical babbling begins around 6 months of age (Stark, 1980). Indeed, in line with a model described previously (Westermann & Miranda, 2004) which involves self-organizing systems, infant researchers have accounted for the fact that infants are gaining experience listening to the language(s) around them at the same time that the human vocal tract is developing (Stark, 1980), which also coincides with the onset of babbling behavior. Thelen (1991) suggests that as infants progress from one babbling phase to another, they receive more and more correlated, stable auditory-proprioceptive feedback, and thus form tighter links between their productive behavior and their auditory environments. This dynamical approach to vocal learning and behavior combines the infants' auditory experiences with the development of the motor (articulatory) system to produce the patterned behavior of speech (Thelen, 1991).

As such, whether an infant's developing vocal production system interfaces with her experience with speech perception is not a new contention (Kuhl & Meltzoff, 1996; Locke, 1990, 2007; Vihman, 1996; Werker & Pegg, 1992). In fact, there is evidence from the vocal development literature to suggest that infant vocal development does interact with speech perception within the first years of an infant's life. For example, infants who are born deaf produce delayed and qualitatively different canonical babbling sounds (Oller & Eilers, 1980), and infants' performance during a speech perception task correlates with language milestones (including word production) in the second year of life (Tsao, Liu, & Kuhl, 2004). As mentioned previously, infants at 4 months of age show evidence of vocal imitation to adults speaking vowel sounds (Kuhl & Meltzoff, 1982; 1996), and some research suggests that preverbal infants' individual babbling patterns influence the way they process speech sounds (DePaolis, Vihman, & Keren-Portnoy, 2011; Majorano, Vihman, & DePaolis, 2014).

However, productive experience is not the *only* kind of sensorimotor or articulatory information that relates to the perception of speech sounds; as seen in the findings from Yeung and Werker (2013), and many of the behavioral studies conducted on adults (Fowler & Dekle, 1991; Gick & Derrick, 2009; Ito et al., 2009), configuration of the articulators (or extra-sensory articulatory information) can affect speech perception (specifically, audiovisual speech perception in infants). Thus, studying infants' overt articulatory or imitative (and later productive) behaviors is not the only way to address the nature of the sensorimotor-auditory link across development.

Taking together the findings from the vocal development and imitation literature cited above, as well as the articulatory-manipulation study implemented in Yeung & Werker, 2013, the available evidence suggests that sensorimotor information interacts with auditory and visual

information during speech perception. Researchers have purported, in fact, that the mapping between auditory and visual speech occurs because of the shared articulatory qualities (Kuhl & Meltzoff, 1982; Yeung & Werker, 2013). Although it seems to be the case that these three systems interact with each other, the origins of the linkage between articulatory information and speech perception may rely on presence of visual speech information. While adults do not require visual access to speech in order for sensorimotor information to affect speech processing (at both the neural and behavioral levels), this possibility has yet to be studied in infants. As it is present in newborn infants, the ability to perceive and integrate auditory and visual speech is either independent of experience or is immediately triggered by the first experience with visual information; the extant evidence on sensorimotor-speech perception links in infancy (which each include visual speech information) seems to suggest that the link between auditory-visualarticulatory information may be experience-independent as well. However, whether the link between sensorimotor information from the articulators and the speech processing system is independent of both a) visual information and b) experience with the native language is an open question.

In order to integrate these theories and the extensive neural and behavioral evidence described previously, and build a framework within which I can investigate the development of sensorimotor-auditory links in speech perception, I now focus on a model that involves dual-stream neural pathways. Specifically, one pathway has been implicated in the integration of sensorimotor and auditory speech information – a *dorsal* stream for language processing.

1.5 Building a framework for development of sensorimotor influences on speech perception: The dorsal stream & multisensory integration in speech processing

Considering the neural areas involved in processing multisensory speech discussed earlier, the pathways and networks that exist between these cortical areas can further elucidate their function during speech perception. Dual-processing pathways during visual perception were discovered in the 1980s: the first, a ventral stream that is involved in the perceptual identification of objects (the 'what' stream), acts as the interface between the visual system and a conceptual system. A second pathway, the dorsal stream, is involved in integrating visual information with the necessary sensorimotor information to guide object directed actions (the 'how' stream); it is in this stream where the visual system interfaces with the motor system (Goodale & Milner, 1992; Mishkin, Ungerleider, & Macko, 1983). In the case of language processing, Hickok and Poeppel (2004, 2007) proposed an analogous dual-stream pathway for the perception of language. Each pathway bridges to and from the Superior Temporal Gyrus (STG); the first is a ventral stream that is involved in conceptual and semantic processing, specifically the phonological-to-lexical representations of speech sounds (the 'what' pathway for language which, they proposed, projects ventro-laterally toward the inferior posterior temporal cortex in the posterior middle temporal gyrus). The second is a dorsal stream that maps the auditory speech representations onto articulatory-motor representations (the 'how' pathway for language which, they proposed, projects dorso-posteriorly through the parietal-temporal boundary (area Spt in the planum temporale), the inferior frontal gyrus (IFG), and the premotor cortex) (Okada & Hickok, 2006; see Figure 1.2 for pictorial representation of pathways, from Friederici, 2011). Saur and colleagues (Saur, Kreher, Schnell, Kummerer, Kellmeyer, Vry Umarova, Musso, Glauche, Abel, Huber, Rijntjes, Hennig, & Weiller, 2008) identified these

streams using fMRI and diffusion tensor imaging (DTI) while adults performed tasks in which they listened to speech. The authors found a ventral stream that connected the middle temporal lobe and the ventrolateral prefrontal cortex, and was involved in higher-level language comprehension, or mapping sound to meaning. The dorsal stream, in contrast, connected the superior temporal lobe and the premotor cortices in the frontal lobe, and was responsible for the sensorimotor mapping of sound to articulation (Saur et al., 2008).





From Friederici, 2011.

A similar model of activation involving two separate streams during multisensory particularly audiovisual—speech perception has been proposed by Okada and Hickok (2009), wherein each stream processes and integrates different forms of speech information to result in a multisensory speech percept. The sensory-sensory (ventral) network integrates auditory and visual speech information (and involves projections from sensory cortices to the STS), while the left-dominant sensory-motor (dorsal) network combines visual speech gestures with the auditory speech information (and involves activation in the posterior inferior frontal gyrus (pIFG), dorsal premotor cortex, the planum temporale (PT), and posterior STS).

1.5.1 Dorsal stream and correlated articulatory-auditory information

Ruth Campbell (2008) further described these ventral and dorsal processing streams and their roles during audiovisual speech processing: in the ventral stream, a *complementary mode* of activation offers 'missing' information to auditory speech during perception, while in the dorsal stream, a *correlated mode* of activation provides redundant or dynamically similar information about the articulatory (motor) behaviors. In the complementary mode (ventral stream), audiovisual speech information is primarily processed along the upper surface of the temporal lobe, and includes the inferior occipito-temporal regions; Bishop and Miller (2009) provide evidence for a similar network in processing audiovisual speech in noise. This ventral stream has been implicated in specifying image details of speech information in the face, including movements of the mouth, lips and tongue, that offer support for auditory speech information; in the complementary mode, the combination of visual and auditory speech can change or improve meaning.

In the correlated mode (dorsal stream), on the other hand, audiovisual speech is generally processed in the superior temporal gyrus to the frontal-temporal-parietal junction (and mainly projects to the pSTS). While visual speech shares many of the dynamic features and patterns with those found in auditory speech, Campbell (2008) argues that in this dorsal stream, the relevant and redundant features from both visual and auditory speech are abstracted and bound into a unified, integrated speech percept.

Whether the ventral and dorsal streams that process auditory (Hickok & Poeppel, 2000; 2004; 2007) or audiovisual speech information (Campbell, 2008; Okada & Hickok, 2009) are

distinct remains an open question, and is outside the scope of this dissertation. However, taking into account the similarities in the areas activated in the dorsal streams during both auditory and audiovisual speech perception (areas including the temporal cortex, the parietal-temporal junction in area Spt, the inferior frontal gyrus, and the premotor cortex), it seems that it is in this stream where the links between auditory and sensorimotor information occur.

Research concerning neural activity and the active networks during audiovisual speech perception in infancy is limited in comparison to the large number of studies conducted with adults. Regarding the patterns of connectivity in activation during audiovisual speech perception, researchers have identified age related differences between adults and children: while both groups exhibit a greater overall BOLD response to audiovisual speech than to auditory speech, there is a significant difference in the connectivity of a pathway in the frontal-temporal-parietal network (from the IFG and ventral premotor cortex, to the supramarginal gyrus (SMG)) (Dick, Solodkin, & Small, 2010), a network very similar to the dorsal, correlated mode proposed by Campbell (2008). Adult networks are more developed for integrating auditory and visual speech information than are the same networks in children (Dick et al., 2010). According to the authors, the functional difference in brain areas between adults and children suggests that, even though the necessary cortical areas are in place much earlier, the development of audiovisual speech

1.5.2 Dorsal stream and language processing in infancy

At what point in development are infants' speech perception systems influenced by the sensorimotor (articulatory system) *without* the addition of visual speech information? This question remains largely unanswered in the literature, both in behavioral and neuroimaging research. However, there is some evidence to suggest that the networks for mapping auditory

speech on to articulatory-motor representations—in a dorsal stream as discussed above (Hickok & Poeppel, 2004, 2007; Friederici, 2011)—is available in newborn infants. Using fMRI (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Dehaene-Lambertz, Hertz-Pannier, Dubois, Meriaux, Roche, Sigman, & Dehaene, 2006) and MEG (Imada, Zhang, Cheour, Taulu, Ahonen, & Kuhl, 2006), research on preverbal infants as young as 3 months of age has identified brain regions that are more active during the presentation of speech sounds over silent periods: the left temporal lobe, the bilateral superior temporal sulci, the left planum temporale, and even the inferior frontal gyrus. By using DTI to identify the pathways between these areas, two distinct language-processing streams have been shown even in newborns: a ventral pathway connecting the inferior frontal gyrus with the superior temporal cortex, and a dorsal pathway connecting the temporal cortex to areas in the premotor cortex (Leroy, Glasel, Dubois, Hertz-Pannier, Thirion, Mangin, & Dehaene-Lambertz, 2011; Perani, Saccuman, Scifo, Anwander, Spada, Baldoli, Poloniato, Lohmann & Friederici, 2011). Later in development, a second dorsal pathway that reaches the inferior frontal gyrus becomes functionally connected to the temporal cortex (see Figure 1.2) (Friederici, 2012). This research suggests, as do the authors, that the existence of these pathways provide preverbal infants with the processing capabilities that integrate sensorimotor information from the articulators with incoming auditory speech information. In fact, recent evidence has shown that experience with the native language selectively affects the pattern of activation in auditory and motor areas between 7 and 12 months of age: in 12-month-olds, non-native sounds activate motor areas to a greater degree than native speech sounds, while native sounds result in greater activation in auditory areas compared to non-native sounds (Kuhl, Ramirez, Bosseler, Lotus Lin, & Imada, 2014).

Considering the evidence for these functional pathways, it seems that some of the necessary streams for processing (sensorimotor influences on) auditory speech are at least *available* in newborns. While further maturation occurs and functional connectivity continues to develop in these streams throughout infancy, the early-existing organization of the ventral and dorsal streams may help to constrain the pathways and circuits that develop with specific multisensory (including articulatory-motor) speech input. If the dorsal pathway is in fact integrating sensorimotor information with auditory speech, one should see early sensorimotor effects during auditory speech perception on the activation patterns in the dorsal network.

Further, models that purport sensorimotor-auditory coupling via productive experience (as in Westermann & Miranda, 2004) seem to be at odds with the early existence of the dorsal pathway for language processing; given the fact that neural architecture is already in place at birth for (what researchers assume should be) linking sensorimotor and auditory speech information, there is a possibility that covert³ movements of the articulators may also interact with auditory speech perception well before infants begin to babble. Across time, such an earlylinkage between sensorimotor and auditory systems may strengthen as an infant gains experience listening to and producing her native language.

To return to Campbell's (2008) idea of processing correlated information in a dorsal stream, it is worth noting that humans are sensitive to and easily process the highly correlated information that exists in natural speech between the auditory and visual domains; the auditory information (auditory envelopes and formant frequencies important for distinguishing speech sounds) and visual structural information (including mouth opening and inter-lip distance) in

³ I use the term 'covert' to mean movements or configurations of the articulators that are *not* self-produced, but extrinsically manipulate the shape of the mouth (including the use of soothers and teething toys).

speech are robustly correlated (Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009). The same is likely true for the correlation between auditory and articulatory information. Thus, it may be the case that, when auditory speech sounds match the qualities and shape of the articulators (even in the absence of vision), an influence of sensorimotor information on infants' perception of auditory speech information may become evident in the early-existing dorsal stream. These unresolved issues lead to the main questions to be addressed in this dissertation:

- 1. Does the sensorimotor-auditory link described previously require experience listening to and seeing speakers produce language? Does it exist at birth?
- 2. Can covert manipulations of the articulators affect auditory speech perception in the first year of life?
- 3. If so, can this early link affect an infants' behavior? Or does it only exist in underlying neural processing, perhaps in areas that overlap with the dorsal stream of language processing?
- 4. Is the sensorimotor-auditory speech perception link one that facilitates or improves speech processing in infants? Or can it also result in inhibitory effects?

1.6 Current experiments

In order to fully understand how the link between the sensorimotor system and the auditory speech perception system develops, and which systems can affect auditory speech processing, more research is necessary. It is possible that the impact of the sensorimotor modalities evident in adult speech processing requires experience producing speech; because preverbal infants have not had experience with speech production, the influence of these sensorimotor modalities during auditory-only speech processing may take time to develop, and may require visual access in order to have an effect. Further, the role of these modalities in infant
speech perception *without* visual speech information has not been empirically addressed, and it could be the case that the sensorimotor modalities affect infants' auditory speech perception abilities before they have experience producing speech themselves.

The goal of this dissertation is to combine the existing knowledge of typical auditory speech perception in infancy with the growing literature on multisensory effects on the perception of speech; specifically, I aim to identify how and if sensorimotor information can affect the perception of speech sounds in infants, without the influence of visual speech information. In the first experimental chapter, I determine whether the sensorimotor system, specifically, the position and movement of the infant's tongue, influences a 6-month-old infant's ability to auditorily perceive two distinct, non-native speech sounds (Experiments 1-4, Chapter 2). In the second experimental chapter, I examine the influence of the sensorimotor system, specifically, the shape and action of the infants' lips, in the learning of and memory for vowel sounds in neonates using the High-Amplitude Suck procedure, a common behavioral technique used to identify perceptual capabilities in neonates and young infants (Experiment 5, Chapter 3). In the third and final experimental chapter, I again study neonatal memory for speech sounds using the High-Amplitude Suck procedure; in addition, I utilize Near-Infrared Spectroscopy (NIRS) to investigate the different neural networks involved in processing sounds that either match or mis-match the infants lip shapes (Experiments 6-7, Chapter 4).

Taken together, these three experimental sets will advance our understanding of infant speech perception by identifying whether speech perception can be facilitated or disrupted by sensorimotor information without the influence of production experience. These theoretically motivated studies with typically developing infants will ultimately help inform intervention in infants who are growing up with speech and language difficulties—including those with

orofacial anomalies and other motor impairments—by pointing to ways in which sensorimotor information might affect or enhance their speech perception and language development.

Chapter 2 : Sensorimotor influences on speech sound discrimination in 6-month-old infants

2.1 Introduction

A continually growing area of speech perception research concerns the multisensory nature of speech, and how the human perceptual system effortlessly integrates speech information across domains. Much of the multisensory speech research to date has concerned audiovisual speech processing; two seminal findings from adult literature exhibit the role of visual speech information during speech perception: first, visual speech information disambiguates auditory speech when it is embedded in white noise (Sumby & Pollack, 1954). Second, visual speech information can change the auditory percept; in an illusion known as the McGurk effect, watching a speaker mouth the sound /ga/ during simultaneous presentation of the auditory sound /ba/ causes the subject to perceive the sound as /da/ (McGurk & MacDonald, 1976). These two pieces of evidence reveal the fact that visual speech augments speech perception in adults.

2.1.1 Multisensory speech perception in infancy

However, in order to accurately characterize how humans represent and process speech and language information, researchers must take into account the development of speech processing, and how speech perception changes across the first months of life. Even in infancy, research has shown that speech perception is also multisensory, as information from extraauditory systems can be integrated with auditory speech information. For example, infants as young as 2 months of age match a speaker's voice to a video of the correct articulating face, as evidenced by their looking behavior (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999, 2003), and integrate auditory and visual speech information, as seen in ERP studies (Bristow et al., 2008; Kushnerenko et al., 2008). Visual speech information also enhances phoneme discrimination in 6-month-old infants (Teinonen et al., 2008). In addition, infants are able to visually discriminate faces articulating their native language from faces articulating an unfamiliar language, suggesting an ability to perceive language information on the basis of visual information alone (Weikum, Vouloumanos, Navarra, Soto-Faraco, Sebastian-Galles, & Werker, 2007). These early effects of audiovisual speech processing in infants suggest that visual speech information may be a privileged form of input in speech processing in infancy.

2.1.2 Sensorimotor influences on speech perception in infancy

While most of the multisensory speech perception studies conducted on infants concern the way visual information influences speech perception, few studies have examined whether sensorimotor information, specifically information from the articulators, affects infant speech perception. As discussed in the previous chapter, in the adult literature, there is much evidence to suggest a modulatory effect of sensorimotor information on speech perception (Hickok et al., 2011; Ito et al., 2009; Sams et al., 2005, Scott et al., 2013). However, the origins of this link in infancy are still largely open for debate, and the development of this link as an infant gains experience with her native language must be taken into account.

In fact, considering for a moment the other side of the production-perception coin, there is evidence from the vocal development literature that speech perception affects production patterns in infancy. For example, performance on speech perception tasks at 6 months of age correlate with language milestones, including word production, at 2 years of age (Tsao et al., 2004). Further, the linguistic environment in which an infant is raised affects the particular pattern of sound production (de Boysson-Bardies & Vihman, 1991); indeed, even the language experienced in utero affects 'production' to some degree, as newborn cries reflect the prosody of

the language spoken by the newborns' mothers in utero (Mampe, Friederici, Christophe, & Wermke, 2009). Infants born with a severe-to-profound hearing impairment exhibit delays in syllable production, and experience a delay in the onset of canonical babbling (Oller & Eilers, 1988). The typical development of speech production involves a sequence of stages (Vihman, 1996): the reflexive stage (0-2 months) is comprised mainly of vegetative sounds and sounds of discomfort and crying. The cooing and laughter stage (2-4 months) is comprised of comfort sounds, and sounds that are quasi-vocalic which may be separated by velar or glottal consonants. The vocal play or expansion stage (4-7 months) is comprised of fully resonant vowels, as well as other sounds that are a product of vocal tract exploration: squeals, whispers, growls and raspberries. During canonical (or reduplicated) babbling (7-10 months), infants start to produce consonant-vowel syllables in repetitive strings, such as /babababa/; it is during this stage that infants begin to produce adult-like phonetic sequences. Finally, the variegated babbling stage (10+ months) is comprised of strings of consonant-vowel syllables with variation in consonant or vowel elements, such as /bagidagodu/. The fact that deaf infants exhibit delays in the stage during which adult-like syllables are produced suggests that a functioning auditory system plays an important role in vocal development. So, does vocal development, or articulatory movement at the very least, affect auditory speech perception in infancy?

As eluded to earlier, the sensorimotor-auditory link could be a result of experience producing speech, meaning sensorimotor-auditory link effects on auditory speech perception would not be fully established until infants have command over speech production. Some research has attempted to address whether individual speech production patterns affect speech perception in infants. DePaolis and colleagues (DePaolis et al., 2011) provide evidence that infants attend more to words that contain sounds which are not in their own production

repertoire, and suggest that babbling affects infants' sensitivity to sounds that are important for early language acquisition (see also Majorano et al., 2014). Further, evidence from individuals with production disorders has shown that children who exhibit production errors (Desjardins, Rogers, & Werker, 1997) and adults with cerebral palsy (Siva, Stevens, Kuhl, & Meltzoff, 1995) have difficulty integrating auditory and visual speech information into a unified percept. Together, these findings do suggest that experience with speech production can influence perception in some situations.

A second possibility is that experience with speech production is not necessarily required for sensorimotor (articulatory) information to affect speech perception, but that having visual access to speech information plays an important role in the sensorimotor-auditory speech link. Specifically, it has been proposed that the information in auditory and visual speech is mapped onto a common articulatory representation (Kuhl & Meltzoff, 1982, 1984; Yeung & Werker, 2013). Researchers first assessed the articulatory-audiovisual speech link by way of imitation studies in infants who are only a few days old. In response to an experimenter producing either /a/ or /m/ sounds, newborns would exhibit more mouth opening after hearing /a/ sounds, and more mouth clutching after hearing /m/ sounds (Chen et al., 2004), and similar imitation patterns were shown for the vowels /a/ and /i/ following audiovisual presentation (Coulon et al., 2013). Further, in Kuhl and Meltzoff's (1982) report of bimodal speech perception in 4-month-olds, they discuss observations of infants producing sounds that resembled vowels in response to seeing and hearing speech stimuli during the task (compared to pure tone, nonlinguistic stimuli) (see also Patterson & Werker, 1999). In follow-up work, infants between 12-20 weeks of age who watched and listened to adults produce vowels 'responded' by producing vowels that perceptually matched those spoken by the adults (Kuhl & Meltzoff, 1996). The authors suggest

that the vocal imitation exhibited by 4-month-old infants reflects infants' recognition (or knowledge) of the intermodal equivalences in the speech information in the auditory, visual, and motor modalities.

In the first empirical test of the role of articulatory information in speech perception, Yeung and Werker (2013) showed that changing the shape of infants' mouths affected their performance in an audiovisual speech matching task; sucking on a soother (forcing lips into a rounded shape) and chewing on a teether (forcing lips to be spread) selectively interacted with 4.5 month-old-infants' perception of audiovisual /u/ or audiovisual /i/ sounds, respectively. These researchers found that infants whose lips were rounded and who were hearing an /u/ sound looked longer to an /i/-articulating face, and that infants whose lips were spread and who were hearing an /i/ sound looked longer to an /u/-articulating face, thus exhibiting an articulatory contrast effect.

Together with the findings from imitation studies, these results suggest that infants are able to integrate auditory and visual speech information with a common sensorimotor (or articulatory) representation, which in the case of Yeung & Werker (2013) caused an interference effect in an audiovisual speech matching task. Therefore, the possible explanation is that sensorimotor information can only interact with speech perception when there is visual speech information available to the infant. Multiple forms of converging speech information from different modalities may be necessary for an infant's perception of speech to be affected by information from her own articulators; only later in development would sensorimotor information interact with auditory speech perception.

In the current experiments, we would like to suggest a third possibility: articulatory information affects speech perception in the absence of perceptual experience⁴ *and* in the absence of visual speech information. We propose that an early link between information from the articulators (in the sensorimotor system) and information in the auditory speech signal (in the speech perception system) exists even without particular auditory or visual experience with a set of speech sounds. To investigate this possibility, we do so by building on the well-documented findings of non-native consonant perception in the first year of life.

2.1.3 Current experiments: Sensorimotor influences on non-native consonant perception in infancy

The speech sound discrimination abilities of preverbal infants are well-documented in the literature, ranging from consonants (Eimas et al., 1971), to vowels (Kuhl et al., 1992), to lexical tones (Mattock, Molnar, Polka, & Burnham, 2008); as discussed in Chapter 1, concerning consonant perception, infants at 6 months of age discriminate both native and non-native speech sounds (e.g. Werker & Tees, 1984). Specifically, 6-month-old English-learning infants are able to discriminate the contrast between a Hindi voiced dental stop consonant (d) and a Hindi voiced retroflex stop consonant (d). The dental /d/ is produced by placing the tongue at the back of the teeth on the roof of the mouth, and the front of the hard palate (see Figure 2.1a), thus in front of the English alveolar /d/. The retroflex /d/, in contrast, is produced behind the English /d/ by curling the tongue tip back behind the alveolar ridge, so that the bottom side of the tongue tip makes contact with the roof of the mouth (see Figure 2.1b).

⁴ We did not take productive inventories of the infants in our studies, therefore, we cannot determine that the infants do *not* have the Hindi /da/ - /da/ distinction in their productive repertoire. However, we controlled for linguistic background, and excluded any infant that was exposed to less than 90% English, ensuring that the perceptual experience did not include the Hindi contrast.

Figure 2.1. Images of Hindi speaker producing the phonemes a) /d/ and b) /d/.





Taking advantage of the fact that 6-month-old English-learning infants are able to discriminate the Hindi $\frac{d}{-d}$ contrast, the experiments in the current chapter investigated whether a functionally available articulatory system is necessary for speech perception. Even though infants at this age are likely not producing the contrasts they hear (they are likely in the vocal play stage of babbling, maybe at the beginning of canonical babbling), the possibility exists that infants recruit information from their own articulators in order to fully perceive the contrast, meaning the necessary articulators must be able to freely move. As discussed above, infants' own articulatory-motor movements (sucking on soother or chewing on a teether) affected the audio-visual perception of vowel sounds (Yeung & Werker, 2013). In the current studies, we used a similar sensorimotor manipulation technique (utilizing teething toys) to investigate whether sensorimotor information could affect the perception of speech sounds in the absence of any visual speech information; in doing so, we used two different teething toys (one that affected tongue tip placement, one that did not affect tongue tip placement) in a non-native speech discrimination paradigm to determine if articulatory information is recruited during speech perception.

2.2 Experiment 1: Non-native speech perception in 6-month-old English-learning infants

The first experiment in Chapter 2 was a control study, in which the non-native speech sound discrimination capabilities of 6-month-old English-learning infants was tested. The contrast used was the Hindi dental /d/ - Hindi retroflex /d/, and each phoneme was embedded in the CV syllables /da/ and /da/. The discrimination of this contrast has been tested many times in 6-month-old infants using the procedure similar to the one used in Experiment 1 (and Experiments 3 & 4), even in our lab (ie, Yeung & Werker, 2009; Weikum et al., 2012). However, because a new speaker was recorded in order to provide higher sound-quality stimuli for the studies in Chapter 2, Experiment 1 was a necessary step to validate the use of the newly-recorded stimuli. The choice for this particular speech sound contrast comes from the fact that the main phonetic difference between these two sounds is in the placement of the tongue during pronunciation. As will be described in Experiments 2 and 3 below, the location of the tongue tip can be disrupted using one of the two teething toys.

2.2.1 Method

2.2.1.1 Participants

The participants in Experiment 1 were twenty-four infants (12 male, 12 female), with a mean age of 6 months, 22 days (ranging from 6 months, 5 days to 7 months, 25 days), with a 97.50 % average exposure to English (ranging from 90-100%English). These infants' parents were contacted through the Infant Studies Centre at the University of British Columbia. Parents and infants were originally recruited for participation at BC Children's and Women's Hospital, after expressing interest in being contacted for research studies. Parents gave written consent for their infant's participation before the study began. After the study, infants received a t-shirt and were awarded a certificate as a thank-you for participating. In addition to the 24 infants included

in the final analyses, data from 3 infants were not included due to fussiness (n = 2) and parental interference (n = 1).

2.2.1.2 Stimuli

The auditory stimuli used in Experiment 1 (and Experiments 3 & 4 below) were recorded by a female native-speaker of Hindi. Stimuli included a set of /da/ and /da/ syllable tokens recorded in infant-directed voice, where each syllable was spoken in a triplet (i.e., /da//da//da/); the middle token was spliced from each triplet, and used in the final audio files. Each syllable token was analyzed for pitch, duration, and amplitude consistency. Three unique tokens of /da/, and three unique tokens of /da/ were used to create the 20 second auditory stimuli streams described below. On average, the tokens were 900 ms in duration and 70 dB in amplitude, and inter-token intervals were between 1000 and 1100 ms in duration.

2.2.1.3 Apparatus

Looking time data were collected using a Tobii 1750 eyetracking system, which consists of a PC-run monitor (34 cm x 27.5 cm screen) that both presented the visual stimuli and captured the infants' gaze information, and a Macintosh desktop that controlled the stimuli presentation. The PC used Tobii Clearview software for the collection of gaze data from the eye-tracking monitor. The eyetracker collected gaze information every 20 ms, and the area of interest for the following analyses included the full-screen (1024 X 768 pixels, or 34 X 27.5 cm). Therefore, during each time interval, looking behavior was classified as either 'looking on screen' or 'no gaze information.' Data from each trial were collected in this way, and looking time was analyzed using a custom Excel script.

2.2.1.4 Procedure

The procedure implemented in Experiment 1 (and in Experiments 3 & 4 below) is a commonly used task to identify speech sound discrimination abilities in infants from 4 - 10 months of age: the alternating/non-alternating sound presentation task (Best & Jones, 1998; Yeung & Werker, 2009; Weikum et al., 2012; Yeung, Chen, & Werker, 2013). In this task, infants experienced two types of trials, one that involved *alternating (A)* speech sounds (including /da/ and /da/), and one that involved repetitions of *non-alternating (NA)* speech sounds (either multiple instances of /da/ or multiple instances of /da/). A total of eight 20-second stimuli streams were created, and each included a series of 10 syllables. In the alternating stimuli streams, the 3/da/and 3/da/tokens were repeated in a pseudo-random order (up to 10)syllables), with the requirement that no one syllable type could repeat more than two times in row. Four alternating streams were created, two of which began with a /da/ token, two of which began with a /da/ token. In the two non-alternating /da/ streams, the 3 tokens of /da/ were repeated (up to 10 syllables), and the same token was never repeated two times in a row. In the two non-alternating /da/ streams, the 3 tokens of /da/ were repeated (up to 10 syllables), and the same token was never repeated two times in a row.

Infants were seated on their parent's lap facing the Tobii 1750 eye-tracking monitor in a dimly lit, sound-attenuated room. Parents were asked not to speak to their infants, and listened

to masking music over headphones. The study began with a calibration period, to ensure proper gaze tracking and recordings for each infant during the study. The experimenter controlled the study at a separate computer in the study room, behind a curtain and out of sight of the infant and parent.

After the evetracker was calibrated, infants were shown a bright, colorful looming ball in the center of the screen to orient their gaze centrally; as soon as the infants looked to the center of the screen, the trials began. During each 20-second trial, a series of 10 sounds were played from two speakers located to the sides of the Tobii 1750 monitor at a level of 65 dB; infants were shown a black and white checkerboard on the monitor during each trial. The purpose of a checkerboard display was to present an image that was unlikely to be treated as an object requiring a label, yet offered high-contrast colors and captured attention. Each trial was a fixed length (20 seconds), and the looming ball 'attention getter' was presented in between each trial to centrally orient the gaze before the next trial began. Infants were assigned to one of 8 unique experimental orders (4 orders began with an alternating trial, and 4 orders began with a nonalternating trial), and every infant experienced all 8 trials. As will be described below, looking time data during each trial were analyzed in *pairs* in order to account for any changes in looking time across the series of trials; for example, pair 1 included the first 2 trials (one alternating, one non-alternating), pair 2 included the third and fourth trials (one alternating, one non-alternating), pair 3 included the fifth and six trials (one alternating, one non-alternating), and pair 4 included the seventh and eighth trials (one alternating, one non-alternating).

2.2.1.5 Hypotheses

Discrimination of the speech sounds was inferred if infants had a preference to look longer at the checkerboard during one of the two types of trials. Typically, if infants are able to

detect a difference between two sounds and there has not been a familiarization phase, they will show longer looking time during *alternating* trials compared to *non-alternating* trials. Importantly, in the alternating/non-alternating procedure, infants' discrimination of sounds is tested—not their learning of the sounds. By including only a "test" phase, the infants' baseline perception of the sounds could be determined, regardless of whether they were able to learn this distinction. In Experiment 1, it was hypothesized that 6-month-old English-learning infants would show longer looking during alternating compared to non-alternating trials, and would thus provide evidence that they successfully discriminate the /d̪a/ - /d̪a/ distinction and replicate previous findings using this contrast.

2.2.2 Results

Looking time data were analyzed across the 4 trials of each type (4 alternating, 4 nonalternating) in the 4 pairs of trials as described earlier; looking times to the alternating and nonalternating trial types were used as the dependent measure in the following analyses. A 2 (Trial Type) X 4 (Pair) repeated-measures ANOVA was performed on the looking times, using the within-subjects factors of Trial Type (alternating or non-alternating) and Pair (1, 2, 3, or 4). There was a significant main effect of Trial Type, F(1,23) = 4.32, p = .049, $\eta_p^2 = .16$, and no interaction with Pair, F(3,69) = 1.63, p = .19, $\eta_p^2 = .066$, suggesting that the pattern of looking time to the two types of trials was similar across the 4 pairs. Follow-up investigation of looking time means for the significant Trial Type effect showed that infants looked longer during alternating trials (M = 9369.17 ms, SD = 4033.06) than non-alternating trials (M = 8542.29 ms, SD = 4053.58), suggesting that infants successfully discriminated the Hindi consonant contrast in this paradigm.

2.2.3 Discussion

The results from Experiment 1 replicated previous findings, showing that 6-month-old English learning infants were able to discriminate the non-native Hindi /d̪a/ - /d̄a/ contrast. Infants showed longer lengths of looking during the alternating trials compared to nonalternating trials. Thus, this justified the use of the same alternating – non-alternating procedure with our newly recorded stimuli in Experiments 3 and 4.

Before turning to the experiments involving the sensorimotor manipulation during the alternating – non-alternating procedure (incorporating teething toys), we wanted to first provide validation for the choice of teethers to be used in Experiments 3 and 4. Following the hypotheses described earlier, if articulatory information is recruited during infant speech perception, a teether that selectively impairs tongue tip movement should also impair discrimination of the Hindi /da/ - /da/ contrast, while a teether that does not affect the tongue tip should not impair discrimination of the contrast. Our goal was to select two teethers: one that selectively impaired the tip of the tongue, and another that did not affect tongue tip placement.

2.3 Experiment 2: Ultrasound images of infant tongue contours

The purpose of Experiment 2 was to validate the choice of teething toys to be used as the sensorimotor manipulations in Experiments 3 and 4. This descriptive study involved a small sample of infants whose tongue movements were recorded using a portable ultrasound machine in 3 situations: a) while they had no teether in the mouth; b) while they had a 'flat' teether (the Learning Curve Baby® Fruity Teether) in the mouth, which was chosen based on the hypothesized impact it would have on the placement of an infant's tongue tip (see Figure 2.2a); and c) while they had a 'gummy' teether (the Nuby Gum-EezTM First TeetherTM) in the mouth, which was chosen based on the hypothesized null-effect it would have on the tongue tip (see

Figure 2.2b). The 'flat' teether, we hypothesized, would impact the tip and blade of the tongue when inserted into the infant's mouth, given its flat, planar shape. The 'gummy' teether was comprised of a soft u-shaped silicone pad which fit between the infant's gums, but due to the u-shape, was not expected to affect the placement of the tongue tip.

Figure 2.2. Images of the a) flat teether and b) gummy teether.



2.3.1 Method

2.3.1.1 Participants

The participants in Experiment 2 were 3 infants (2 male, 1 female), with a mean age of 7 months, 10 days (ranging from 6 months, 23 days to 7 months, 25 days). These infants' parents were contacted through the Infant Studies Centre at the University of British Columbia. Parents and infants were originally recruited for participation at BC Children's and Women's Hospital, after expressing interest in being contacted for research studies. Parents gave written consent for their infant's participation before the study began. After the study, infants received a t-shirt and were awarded a certificate as a thank-you for participating.

2.3.1.2 Apparatus

The equipment used to record ultrasound images included a portable ultrasound machine, the 2 teething toys ('flat' and 'gummy' teethers), and video recording software. We used a Sonosite Titan portable ultrasound machine, and a 5–8 MHz Sonosite C-11 transducer with a 90° field of view and a depth of 8.2 cm. All recordings were done using the Pen setting. Ultrasound videos were captured on a separate PC computer using a Canopus TwinPact100 converter from the Sonosite ultrasound machine. Twelve separate frames from each of the three teether conditions (from each infant) were spliced from the video recordings into jpeg images. EdgeTrak software (Li, Kambhamettu, & Stone, 2005) was then used to extract the (x,y) coordinates of the infants' tongue contours; 30 coordinate points were extracted for each tongue curve. Statistical analyses (smoothing spline ANOVA, Davidson, 2006) were performed on these tongue contours separately for each infant, as described below.

2.3.1.3 Procedure

Each infant was seated on his/her parent's lap in a comfortable, upright position. No stimuli were presented during the ultrasound sessions, but the parents and experimenters spoke to each other, and to the infant, during the session. The experimenter placed the transducer under the infant's chin, until proper placement was achieved—in which the tongue tip was visible on the Sonosite screen. As soon as the infant was comfortable with the transducer, a second experimenter began the video recording. The experimenter held the transducer in place for 10-20 seconds, or as long as the infant remained comfortable, to obtain an ultrasound recording of the tongue contour with no teether in the mouth.

The parent then placed one of the two teethers into the infant's mouth, and allowed the infant to become comfortable with the teether. As soon as the infant was comfortable, the experimenter again placed the transducer under the chin, and held it there until 10-20 seconds of ultrasound recording had passed (see Figure 2.3 for an image of an infant in ultrasound recording with the flat teether in the mouth). Finally, the second teether was placed in the mouth and ultrasound images were again recorded for a period of 10-20 seconds. The ultrasound recording

sessions lasted different lengths of time, depending on the cooperation of the infant, and ease of

transducer placement.

Figure 2.3. Ultrasound image from live recording (flat teether in the mouth).

The tongue contour ('profile' of the infant's tongue) is the thick, curved white line in the middle of the sonogram; the tongue tip is to the left, the tongue base is to the right. At the tongue tip, notice the indent in the tongue contour due to placement of the flat teether.



2.3.1.4 Data analysis & hypotheses

As mentioned previously, 12 images from each teether condition were spliced from the video recordings; images were chosen based on the clarity of the image and tongue placement (see Figure 2.3 for an example ultrasound image). Smoothing spline (SS) ANOVAs (Davidson, 2006) were used to analyze the ultrasound images collected for the 3 infants in custom R scripts; SS ANOVAs are used to determine whether the shapes of multiple curves (in this case, curves that represent tongue contours during the 3 teether conditions) are significantly different from each other, and are thus ideal for analyzing ultrasound data for individual participants. SS ANOVAs do not return F values, unlike standard ANOVAs; instead, 95% Bayesian confidence intervals were constructed to specify whether and where two curves differ from each other. If the confidence intervals overlapped between any of the two teether conditions, there were not

significant differences in tongue contours between those two conditions. Because we were most interested in the placement of the tongue tip, the area of interest for the following SS ANOVA analyses concerned only the front portion of the tongue, which is always the left side of the tongue contour (the tongue tip/blade).

2.3.2 Results

The three infants' SS ANOVA plots are shown below, separated in pairs of three teether conditions (Figures 2.4 - 2.6). The tongue contours depicted in each figure show the shape of the infant's tongue while chewing on the different teethers, or chewing on no teether; the left end of each tongue contour represents the tongue tip, and the right end of the contour represents the tongue root. As can be seen in Figure 2.4, in all 3 infants, there was no significant difference in tongue tip placement between the no teether (gray lines) and gummy teether (teal lines) conditions; however, as depicted in Figure 2.5, there were differences in tongue tip/blade contours between the no teether and flat teether (pink lines) conditions, as well as between the gummy teether and flat teether conditions (see Figure 2.6).



Figure 2.4. Tongue contours of 3 infants comparing gummy teether to no teether. Dotted lines denote 95% CIs.



Figure 2.5. Tongue contours of 3 infants comparing flat teether to no teether. Dotted lines denotes 95% CIs.



Figure 2.6. Tongue contours of 3 infants comparing flat teether to gummy teether. Dotted lines denote 95% CIs.

As seen in Figure 2.4, 95% confidence intervals overlap in the area of the tongue tip when the three infants chewed on the gummy teether compared to having no teether in the mouth. In contrast, as seen in Figure 2.5, when the infants chewed on the flat teether, the tongue tips/blades were pushed downwards relative to having no teether in the mouth, as the 95% confidence intervals do not overlap in the area of the tongue tip. Likewise, as shown in Figure 2.6, 95% confidence intervals do not overlap between the tongue contours while chewing on the flat teether compared to the gummy teether. Although the particular shape, contour and size of the tongues differed across infants, the important thing to note is that the *flat* teether affected the shape of the anterior portion of the tongue in every infant, either compressing it from the tip to the blade (Subject 05), forcing it downwards at the tip (Subject 13), or pushing it downwards at the tip and back into the mouth (Subject 14); thus, the flat teether prevented contact between the tongue apex and the alveolar or post-alveolar ridge in the three infants. The gummy teether, on the other hand, did not significantly affect the placement or shape of the tongue tip (Figure 2.4).

2.3.3 Discussion

Given the ultrasound images from the three infants tested here, we showed support for the choice of teething toys to be used in the following two experiments. The flat teether significantly impacted the infants' tongue tips, while the gummy teether had minimal to no effect on the infants' tongue tips (compared to having no teether in the mouth). Thus, we find that the tongue tip can be 'selectively' impaired when an infant chews on the flat teether, while the tongue tip remains largely unimpaired when an infant chews on the gummy teether.

2.4 Experiment 3: Non-native speech perception by English-learning infants with an articulatory-motor perturbation

The third experiment in Chapter 2 investigated whether preverbal infants would fail to discriminate the Hindi dental-retroflex contrast during the selective impairment of their tongue tips (using a flat teether). Following the design in Experiment 1 above, Experiment 3 also employed the alternating/non-alternating test procedure, during which infants chewed on a teether that impeded the movement or placement of the tongue tip. The teethers used in Experiment 3 were the Learning Curve Baby® Fruity Teethers; these teethers are flat, wedge-shaped plastic teething toys. Importantly, when infants chew on these teethers, they tend to disrupt the placement of the tongue tip, as discussed in Experiment 2.

With this manipulation, we identified whether temporarily disabling the tongue's movement would affect auditory perception of speech sounds that, in order to properly produce, require a distinction in tongue placement. Although infants at this age do not consistently and purposefully produce these sounds, it is possible that temporarily disabling the necessary articulator may affect the ability to discriminate the fine articulatory distinction in this contrast.

2.4.1 Method

2.4.1.1 Participants

The participants in Experiment 3 were twenty-four infants (12 male, 12 female), with a mean age of 6 months, 30 days (ranging from 6 months, 7 days to 7 months, 25 days), with a 98.92 % average exposure to English (ranging from 90-100%English). In addition to the 24 infants included in the final analyses, data from 3 infants were not included due to fussiness (n = 1), equipment failure (n = 1), and parental interference, including dropping the teether (n = 1).

Parents gave written consent for their infant's participation before the study began. After the study, infants received a t-shirt and were awarded a certificate as a thank-you for participating.

2.4.1.2 Stimuli & apparatus

The stimuli and apparatus used in Experiment 3 were identical to those used in Experiment 1. The teething toy used in Experiment 3 was the 'flat' Learning Curve Baby® Fruity Teether.

2.4.1.3 Procedure

Experiment 3 used the same experimental procedure as Experiment 1 above, with one important addition: 6-month-olds tested in Experiment 3 chewed on a flat, soft plastic teether during the study, which flattened the tongue and impeded its contact with the roof of the mouth (see Figure 2.2a). The teethers used in Experiment 3 were the Learning Curve Baby® Fruity Teethers, referred to as the 'flat' teether in Experiment 2. As in Yeung and Werker (2013), the teether was held in place by the caregiver for the entire study period. Infants' level of preoccupation with the teether was coded offline by a blind coder after the study; infants whose parents dropped the teether at any point during the study were excluded from the final sample (n = 1). As in Experiment 1, 6-month-old English-learning infants experienced 8 trials of either alternating or non-alternating /d̪a/ or /d̥a/ syllables, and their looking time to a checkerboard projected onto a computer screen was recorded using a Tobii 1750 eyetracker. Infants in Experiment 3 experienced these 8 trials while the caregiver held the teether in place.

2.4.1.4 Hypotheses

If infants recruit information from their own articulators while perceiving speech sounds—even those that they've likely never heard before—it was hypothesized that infants in

Experiment 3 would not show any difference in looking time between alternating and nonalternating trials. This pattern of results would demonstrate a *failure* to discriminate the normally-perceivable speech sound distinction (as seen in Experiment 1) when movement of the tongue is temporarily impaired.

2.4.2 Results

As in Experiment 1 above, looking time data were analyzed across the 4 trials of each type (4 alternating, 4 non-alternating) in the 4 pairs of trials as described above; these looking times to the alternating and non-alternating trial types were used as the dependent measure in the following analyses. A 2 (Trial Type) X 4 (Pair) repeated-measures ANOVA was performed on the looking times, using the within-subjects factors of Trial Type (alternating or non-alternating) and Pair (1, 2, 3, or 4). There was no main effect of Trial Type, F(1,23) = 0.011, p = .92, $\eta_p^2 = .001$, and no interaction with Pair, F(3,69) = 1.17, p = .33, $\eta_p^2 = .048$, suggesting that the pattern of looking time to the two types of trials was similar across the 4 pairs of trials. Across the 4 trial pairs, infants did not look any longer during alternating trials (M = 9846.67 ms, SD = 4099.99) compared to non-alternating trials (M = 9803.54 ms, SD = 4311.07) (see Figure 2.7), suggesting that infants failed to discriminate the Hindi contrast while chewing on the teether.

Figure 2.7. Experiment 3 looking time averages during test trials.

a) Average looking time to alternating and non-alternating trials across all 4 pairs. b) Average looking time to each type of trial (alternating and non-alternating) across the 4 pairs of trials. Error bars denote standard errors of the mean.



To ensure that this failure to discriminate the contrast was not due to the level of distraction the teether may have caused the infants, each video was offline coded for the overall level of distractibility by a coder who was blind to the experiment's hypotheses; on a scale of 1 (very distracted) to 7 (not at all distracted), infants were given a global rating of distraction by the teether—taking into account how often they grabbed it, took it out of their own mouths, looked at it, and played with it. Infants received an average rating of 4.92 (SD = 1.59), indicating

that on average, the teether was only mildly distracting to the infants. To investigate whether this level of distraction interacted with performance in the looking time task, infants' teether distraction scores were split into two levels – those who scored 1-4 (highly to mildly distracted, n = 9), and those who scored 5-7 (little to no distraction, n = 15) – which were used in the Teether Distraction between subjects factor in a second ANOVA. A 2 (Trial Type) X 4 (Pair) X 2 (Teether Distraction) mixed ANOVA showed no main effect of Trial Type, F(1,22) = .009, p =.92, $\eta_p^2 = .001$; no interaction between Trial Type and Teether Distraction, F(1,22) = .001, p =.99, $\eta_p^2 = .001$, and no interaction between Trial Type, Teether Distraction, and Pair, F(3,66) =1.65, p = .19, $\eta_p^2 = .070$. This ANOVA suggests that the teether did not affect looking times due to its distractibility, adding support to the conclusion that by selectively impairing the important articulator, the Hindi contrast became more difficult for the infants to discriminate in the alternating/non-alternating task.

2.4.3 Discussion

The results from Experiment 3 provide evidence that sensorimotor information influences speech perception of non-native speech sounds in 6-month-old infants. The fact that infants who chewed on the flat teether and were temporarily inhibiting movement of their tongue tips (as seen in Experiment 2) had difficulty discriminating the Hindi contrast suggests that infants do recruit information from their own articulators during the perception of auditory speech information. These findings add to the existing evidence that articulatory-motor information affects audiovisual speech perception (Yeung & Werker, 2013), and that it does so even in the absence of visual speech information. Because we chose to use a non-native contrast, we suggest that these findings are independent of experience listening to the sound contrast—and likely the experience of attempting to produce or imitate these sounds. Further research is

necessary to validate this last claim; production and imitation data were not collected or included in these studies.

Although the level of distraction by the teether did not interact with looking times for the two trial types, it could be the case that having any teether in the mouth inhibits discrimination of this contrast. Rather than the null effect of looking time between the two trial types resulting from the selective impairment of the related articulator, it could be that chewing on any kind of teething toy affects speech discrimination in a looking time task. Experiment 4 investigates this possibility.

2.5 Experiment 4: Non-native speech perception in English-learning infants with a nonspecific articulatory-motor perturbation

The final experiment in Chapter 2 is a control experiment for the teether effect in Experiment 3; because infants in Experiment 3 failed to discriminate the contrast while chewing on a teether that impeded tongue movement, it may have been due to the fact that the teether was generally disruptive, or drew attention away from the task. Experiment 4 once again employed the alternating/non-alternating test procedure as well as the use of a teether, but here the teether used did not impede the movement or placement of the tongue tip. The teethers used in Experiment 4 were Nuby Gum-EezTM First TeethersTM; these teethers are made of a soft u-shaped silicone material with a pacifier-style back. Importantly, when infants chew on these teethers, the teethers do not largely disrupt the placement of the tongue, as seen in Experiment 2 above.

2.5.1 Method

2.5.1.1 Participants

The participants in Experiment 4 were twenty-four infants (12 male, 12 female), with a mean age of 6 months, 27 days (ranging from 6 months, 6 days to 7 months, 25 days), with 97.63% average exposure to English (ranging from 90-100% English). In addition to the 24 infants included in the final analyses, data from 8 infants were not included due to fussiness (n = 3), equipment failure (n = 2), and parental interference, including dropping the teether (n = 2) and talking during the study (n = 1). Parents gave written consent for their infant's participation before the study began. After the study, infants received a t-shirt and were awarded a certificate as a thank-you for participating.

2.5.1.2 Stimuli & apparatus

The stimuli and apparatus used in Experiment 4 were identical to those used in Experiments 1 and 3. The teething toy used in Experiment 4 was the 'gummy' Nuby Gum-Eez[™] First Teether[™].

2.5.1.3 Procedure

As in Experiments 1 and 3, 6-month-old English-learning infants experienced 8 trials, involving either alternating or non-alternating /da/ - /da/ syllables, and their looking time to a checkerboard was the dependent measure of interest as measured by a Tobii 1750 eyetracker. During all trials, the caregiver held the Nuby Gum-EezTM First TeethersTM (referred to as the 'gummy' teether in Experiment 2) in the infant's mouth (see Figure 2.2b). Infants' level of preoccupation was coded offline by a blind coder after the study; infants whose parents dropped the teether at any point during the study were excluded from the final sample (n = 2).

2.5.1.4 Hypotheses

It was hypothesized that infants in Experiment 4 would exhibit longer looking to the alternating compared to the non-alternating test trials, showing *successful* discrimination of the /da/ - /da/ contrast. Importantly, by allowing the infant to chew on a teething toy that does *not* impede the tongue tip, we expected that this 'gummy' teether would not affect discrimination of the /da/ - /da/ contrast, and would thus indicate a specificity of the temporary motor impairment during speech perception (as found in Experiment 3). Further, successful discrimination would decrease the possibility that the hypothesized effect in Experiment 3 is due to a disturbance or decrease in attention to the task.

2.5.2 Results

As in the analyses conducted in Experiments 1 and 3, looking time data were analyzed across the 4 trials of each type (4 alternating, 4 non-alternating) in the 4 pairs of trials as described above; looking times to the alternating and non-alternating trial types were used as the dependent measure in the following analyses. A 2 (Trial Type) X 4 (Pair) repeated-measures ANOVA was performed on the looking times, using the within-subjects factors of Trial Type (alternating or non-alternating) and Pair (1, 2, 3, or 4). There was a significant main effect of Trial Type, F(1,23) = 5.26, p = .031, $\eta_p^2 = .19$, and no interaction with Pair, F(3,69) = .37 p = .77, $\eta_p^2 = .016$, suggesting that the pattern of looking time to the two types of trials was similar across the 4 pairs. Across the 4 pairs of trials, infants looked longer during alternating trials (M = 10986.46 ms, SD = 4778.55) compared to non-alternating trials (M = 10324.38 ms, SD = 4640.78) (see Figure 2.8), suggesting that infants showed successful discrimination of the Hindi contrast while chewing on the gummy teether.

Figure 2.8. Experiment 4 looking time averages during test trials.

a) Average looking time to alternating and non-alternating trials across all 4 pairs. b) Average looking time to each type of trial (alternating and non-alternating) across the 4 pairs of trials. Error bars denote standard errors of the mean. * indicates significance at p < .05.



As in Experiment 3 above, each infant video was offline coded by a blind coder for overall level of distractibility by the gummy teether; on a scale of 1 (very distracted) to 7 (not at all distracted), infants were given a global rating of distraction by the teether. Infants in Experiment 4 received an average rating of 5.75 (SD = 1.48), indicating that on average, the teether was again only mildly distracting to the infants. An independent samples t-test

comparing the distraction scores from Experiment 3 (flat teether, M = 4.92) and Experiment 4 (gummy teether, M = 5.75) indicated a marginally significant difference in distraction by the teethers, t(46) = -1.88, p = .066, 95%CI for the difference [-1.72, .06]. This comparison suggests that the flat teether is only slightly more distracting on average compared to the gummy teether.

To investigate whether the level of distraction by the gummy teether in Experiment 4 interacted with performance in the looking time task, infants' teether distraction scores were split into two levels – those who scored 1-4 (highly to mildly distracted, n = 6), and those who scored 5-7 (little to no distraction, n = 18)—which were used in the Teether Distraction between-subjects factor in a second ANOVA. A 2 (Trial Type) X 4 (Pair) X 2 (Teether Distraction) mixed ANOVA showed a main effect of Trial Type, F(1,22) = 6.74, p = .016, $\eta_p^2 = .24$; no interaction between Trial Type and Teether Distraction, F(1,22) = 1.41, p = .25, $\eta_p^2 = .060$, and no interaction between Trial Type, Teether Distraction, and Pair, F(3,66) = .94, p = .43, $\eta_p^2 = .041$. As seen in Experiment 3, this follow-up ANOVA suggests that the teether did not interact with looking times due to its distractibility; regardless of level of distraction, infants in Experiment 4 successfully discriminated the Hindi contrast.

2.5.2.1 Results: Comparison of Experiment 3 and Experiment 4

Because the experiences of the infants in Experiments 3 and 4 were hypothesized to be qualitatively different from one another, each study was described and analyzed as a separate experiment. The idea that the flat teether makes perception of a consonant contrast more difficult than a non-impeding gummy teether during the entire duration of the behavioral paradigm warrants this decision, as these two studies were conceptualized to be more distinct than a minor change in condition. Further, although we find significant differences in looking time to the alternating over the non-alternating trials in Experiments 1 and 4 (seen in the ANOVAs and in the effect sizes) but not in Experiment 3, the mean differences in looking time are quite small, even in Experiments 1 and 4. Therefore, we did not expect to find a significant interaction between trial type and experiment in an analysis that combined Experiments 3 and 4 in a single mixed ANOVA.

However, a final analysis was conducted to directly compare the looking time results between Experiments 3 and 4 to investigate whether a statistically significant interaction existed between the looking time patterns in Experiments 3 and 4; a 2 (Experiment) X 2 (Trial Type) X 4 (Pair) mixed ANOVA was performed on the looking times, using the between subjects factor of Experiment (flat teether—Experiment 3 or gummy teether—Experiment 4), and the withinsubjects factors of Trial Type (alternating or non-alternating) and Pair (1, 2, 3, or 4). As hypothesized, there was no significant interaction between Experiment and Trial Type, F(1,46) =1.52, p = .22, $\eta_p^2 = .032$, nor between Experiment, Trial Type, and Pair, F(3,138) = 1.18, p = .32, $\eta_p^2 = .025$, suggesting that looking time patterns to the alternating and non-alternating trials did not differ between the two studies (or across the four pairs) (See Figure 2.9). Interestingly, we also failed to find a main effect of Trial Type, F(1,46) = 1.97, p = .17, $\eta_p^2 = .041$; this suggests that the significant difference in looking time to alternating and non-alternating trials found in Experiment 4 was not detectable when combined with looking time data from Experiment 3.



Figure 2.9. Average looking times during test trials for Experiments 3 and 4. Error bars represent standard errors of the mean.

2.5.3 Discussion

The results from Experiment 4 suggest that teething toys do not generally disrupt performance on the non-native speech sound discrimination task; instead, while chewing on a gummy teether that did not impede tongue movement, 6-month-old English learning infants successfully discriminated the Hindi /da/ - /da/ contrast. Further, these findings provide corroborating evidence that the teether effect seen in Experiment 3 is due to the particular (shape of) teether used, which selectively impacted an infant's tongue tip and blade. When the part of the articulatory system that is necessary to produce adult-like /da/ - /da/ syllables was *not* impaired, even if other parts of the mouth were occupied, infants showed longer looking to alternating compared to non-alternating trials, as in Experiment 1.

2.6 General discussion

In three looking-time experiments, supported by a descriptive analysis of tongue contours using ultrasound technology, this chapter showed that 6-month-old English-learning infants' speech sound discrimination abilities interact with the selective inhibition of related articulatorymotor movements. Given the critical findings from Experiment 3, we show that a flat teether which temporarily impaired tongue movement also impaired the discrimination of a speech sound contrast; therefore, we suggest that preverbal infants do recruit information from their articulators in order to process speech information. Importantly, the articulatory-induced inhibitory effect was evident in infants who have had no experience listening to the particular Hindi non-native contrast, and without concurrent visual speech information. This raises the possibility that fully functional articulatory systems are required for the development of speech perception, or, at the very least, that articulatory information which 'contradicts' the speech sounds an infant is hearing makes some speech sound distinctions more difficult to perceive. Here, we find an inhibitory effect, in which a consonant speech sound contrast becomes more difficult to perceive when competing articulatory information is present during the task; instead of using teethers which matched one of the two speech sounds, we chose to utilize a teething toy that would collapse the $\frac{da}{-}$ $\frac{da}{-}$ contrast into a single percept. Further research is necessary to determine how the infants categorize the $\frac{d}{-}$ + $\frac{d}{-}$ phonemes during articulatory inhibition, and whether they are in fact temporarily collapsing the two sounds into a single category, perhaps into the English /d/.

2.6.1 Will this change across development, and in concert with other sensory systems?

We provide evidence that neither a) experience with a particular speech sound contrast, nor b) visual speech information is necessary for articulatory information to affect speech processing in infants. However, these are not mutually exclusive events—the relationship between an infant's developing speech perception system (which integrates auditory and visual speech information), and the production (articulatory) system is complex, as the two develop in
the same period of time. Further research is necessary to determine if and how speech information (both auditory and audiovisual) interacts with articulatory information as infants become native speech perceivers, and increasingly proficient speech producers. In the first year of life, infants' perceptual systems become tuned to their language environment (Kuhl et al, 1992; Pons et al., 2009; Werker & Tees, 1984) and their early production patterns reflect the ambient language (De Boysson-Bardies & Vihman, 1991). Thus, future studies must begin to untangle the complex interplay between these processes, and take into account the information available to infants in their articulators, as sensorimotor information seems to be an important factor in infant speech perception capabilities.

2.6.2 Domain-general or domain-specific: Is this special to speech?

As is the case with much of the work in speech perception literature, it must be determined whether the sensorimotor-auditory links shown in these experiments are special to speech processing, or if these effects are domain-general. There is much evidence to suggest privileged processing to linguistic stimuli across development (Vouloumanos & Werker, 2004, 2007; Vouloumanos, Hauser, Werker, & Martin, 2010). In audiovisual speech perception studies in both infants (Kuhl, Williams, & Meltzoff, 1991) and adults (Tuomainen et al., 2005), visual speech information is not integrated with spectrally-similar, non-speech stimuli, suggesting that multisensory integration of speech information is domain-specific (Vatakis, Ghazanfar, & Spence, 2008).

However, we specifically used a non-speech articulatory manipulation (by utilizing teething toys), so further discussion is necessary. On the one hand, we do not think that the flat teether made the phoneme distinction more difficult to discriminate simply because there was invariable information (in the articulators) available during the presentation of the two speech

sounds; if that were the case, the gummy teether (Experiment 4) would have also resulted in a failure to discriminate the contrast. We purport that the interference is selective, and that inhibiting the tongue tip is the important factor in the results shown in Experiment 3.

Follow-up studies can address this question in two ways: the first is by investigating whether the inhibiting effect of the flat teether would be seen during discrimination tasks that involve non-speech stimuli, such as tones (or sine-wave speech). Given the teether distraction data reported in Experiment 3 (which showed that the infants' level of distraction did not interact with their performance in the task), we hypothesize that the flat teether effect we found is specific to the speech sound contrast, and that it would not impede the discrimination of non-speech stimuli. The second way to address the speech-specific question is to ask whether the flat teether affects the perception of other (non-native) speech sound contrast; we argue that the lack of discrimination is due to the selective inhibition of the articulator that is required for the production of the /da/ and /da/ syllables. Further research is necessary to verify that inhibiting the tongue tip does not interfere with speech sounds that do not involve tongue tip movement, such as the Arabic /k/-/q/ contrast, which involves the back of the tongue and the velum for production.

2.6.3 Conclusions

The findings from Experiments 1 - 4 suggest that sensorimotor information from the articulators selectively affects speech perception in 6-month-old infants regardless of productive or visual experience with the speech sounds. These results extend previous research concerning the interaction between articulatory information and the (audiovisual) processing of speech: they suggest that a link between the articulatory and speech perception systems in infants may be more direct than previously thought, and may be available even before infants accrue experience

producing speech sounds themselves. We argue that theories of infant speech perception must continue (or even begin) to account for the seemingly important role of sensorimotor information as infants become native language perceivers.

Chapter 3 : Sensorimotor influences on memory for speech sounds in newborns

3.1 Introduction

3.1.1 Neonatal speech perception

At birth, humans are sensitive to many aspects of language. They have general preferences for linguistic input, including their own mothers' voice over the voice of another female (DeCasper & Fifer, 1980), their native language over a rhythmically dissimilar language (Byers-Heinlein et al., 2010; Mehler et al., 1988), and speech over non-speech (Vouloumanos & Werker, 2007a). Further, as mentioned in Chapter 1, newborns are also able to process more particular aspects of linguistic information. They exhibit memory for a passage read to them while in utero (DeCasper & Spence, 1986), detect changes in the sounds of a series of syllables (Bertoncini et al., 1988), and remember words to which they were habituated after a 24-hour period of time (Swain, Zelazo, & Clifton, 1993). Recent evidence suggests that newborns respond differently to individual sounds based on the language spoken by the mother while in utero; newborn infants show a greater sucking response in order to elicit prototypical vowels from an unfamiliar, non-native language that was not spoken by the mother while in utero, compared to familiar, native vowels (Moon, Lagercrantz, & Kuhl, 2013). Together, these findings suggest that newborns are already tuned to language.

By studying perception in neonates, researchers are presented with a time period during which the extra-utero environment is novel; neonates have only had a short period of exposure to full-spectrum sounds, light, and tactile experiences. While it is clear that newborns have already had experience with auditory information (the auditory system is functional at 26 weeks gestation, Eisenberg, 1976; Fifer & Moon, 1990; Graven & Browne, 2008; Moore & Linthicum,

2007), as well as tactile and sensorimotor experiences (e.g., fetuses exhibit suckling and swallowing behaviors by 15-18 weeks gestation, Miller, Sonies, & Macedonia, 2003) in utero, the kinds of stimuli available to infants after birth are qualitatively different. Studying the way newborn infants combine and perceive multiple types of information (including auditory, visual, and sensorimotor) across domains would provide insight to the early biases or capabilities for multisensory processing in humans.

In the realm of speech perception research, there is much evidence to suggest that speech processing is multisensory in even the youngest infants (Soto-Faraco et al., 2012). In processing audiovisual information, for example, newborns look longer to a face whose articulators match the sound they are hearing, compared to a face articulating a different sound (mismatching face) (Aldridge, Braga, Walton, & Bower, 1999). Audiovisual speech perception is evident in older infants as well (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999, 2003); however, Aldridge and colleagues (1999) argue that the presence of intermodal matching in newborns provides evidence that the multisensory perception of speech is not due to learning or extensive experience with language.

Concerning proprioceptive or sensorimotor processing, one of the earliest pieces of evidence that newborns represent perceptual experiences across domains comes from imitation work. Shortly after birth, neonates imitate facial gestures displayed by adult experimenters, including tongue protrusion, mouth opening, and lip protrusion (Meltzoff & Moore, 1977). The authors argue that these behaviors are due to infants' ability to represent the intermodal information common to both the visual and proprioceptive modalities, by way of an abstract representational system and a process of Active Intermodal Mapping (AIM; Meltzoff & Moore, 1989) rather than an innate releasing mechanism (Tinbergen, 1951). Applying this intermodal

mapping to speech processing, recent evidence suggests that newborn infants exhibit imitationtype behaviors across modalities while processing speech sounds: neonates showed more mouth opening in response to an experimenter producing an open vowel /a/, and more mouth closing in response to an experimenter producing the consonant /m/ (Chen et al., 2004). While these authors suggest that newborns were mapping the information common to the auditory signal and the infants' own articulatory systems, many of the infants' eyes were open during the studymaking it highly likely that they were also using the visual information exhibited by the experimenter while processing and imitating the articulation of the speech sounds; indeed, neonates are able to imitate the mouth movements of the vowels /a/ and /i/ in response to audiovisual presentation of the sounds (Coulon et al., 2013). Together, it seems the available evidence for AIM models of newborn perception and its role in the sensorimotor or proprioceptive system involve the visual system in some way. In the current work, we are not coming from an imitation point of view per se; however, it is worth noting that the link between the sensorimotor system and the auditory (linguistic) system without any visual information remains unstudied in newborns.

As discussed in Chapter 1, there is plentiful evidence for a link between the sensorimotor and speech processing systems in adults; this linkage seems to be a modulatory one, in that recruitment of information from the articulatory-motor system is not required for the perception of speech (Hickok et al., 2011). If we make the assumption that the shape of the mouth and articulatory system could impact the perception of speech as it does in adults, one should expect there to be a link between mouth shapes and *infant* speech perception as well; indeed, 4-monthold infants performed differently in an audiovisual speech matching task when their mouth shapes were controlled by either sucking on a soother (rounded lips) or chewing on a pacifier

(spread lips) (Yeung & Werker, 2013). This articulatory-motor effect on speech perception remains unstudied in infants younger than 4 months of age (see Chapter 2 for evidence in 6-month-olds). If experience listening to speech and language is required for this link to develop, sensorimotor effects on speech processing should not be evident until later in development, and certainly not at birth. In order to test whether an early link between information from the sensorimotor system and the speech processing system exists, we manipulated the shape of newborn infants' mouths using a well-known technique in neonatal perception—the high amplitude suck procedure (HAS)—causing infants to have rounded lip shapes.

3.1.2 High Amplitude Suck procedure

High Amplitude Sucking (HAS) is a powerful behavioral technique used to investigate early learning, preference, and discrimination capabilities in newborn infants, as newborns are able to readily learn contingencies between their sucking behavior and a contingently presented stimulus which acts as a reinforcer (Byers-Heinlein et al., 2010; DeCasper & Fifer, 1980; Eimas et al., 1971; Floccia, Christophe, & Bertoncini, 1997; Mehler et al., 1998; Moon et al., 2013; Shi, Werker, & Morgan, 1999; Vouloumanos & Werker, 2007a). A common interpretation of the HAS procedure is that it taps into contingency learning in the infant (Floccia et al., 1997). For example, if an infant is reinforced with a sound each time she delivers a high amplitude suck, the infant can learn the relationship between her behavior and a sound reward. In this case, an initial increase in sucking typically occurs, after which the reinforcing properties wear off and sucking rates will decrease (Jusczyk, 1985). Learning and memory can then be tested by introducing either a novel or a familiar sound to the infant—if she 'remembers' the initial sound, sucking rates will increase only to a novel sound stimulus, not to a familiar one. This behavioral measure allows researchers to investigate some of the earliest learning capabilities in newborn infants. While contingency learning in infancy has a long history in theories of infant-parental attachment (see Watson, 2001) and how behavior is modified after contingent presentation of a particular stimulus, as discussed above, it is also used to study perceptual processing in infants using the high-amplitude suck procedure. The nature of the 'reward' stimuli used in high-amplitude suck procedures have taken into account ecologically relevant stimuli—including intrauterine heartbeat sounds (DeCasper & Sigafoos, 1983) and the sound of a mother's voice (DeCasper & Fifer, 1980)—as well as more general linguistic stimuli described earlier. To borrow an idea from John Watson concerning 'contingency awareness', if

"...an organism's functional knowledge that the nature of the stimuli received is sometimes affected by the nature of the behavior the organism is emitting" (Watson, 1966, page 123),

then perhaps an infant learning to associate her own suck behavior with a particular resultant sound will exhibit differential learning compared to an infant who receives a different resultant sound. In the context of HAS, in which the behavior of an infant involves sucking on a pacifier that forces the mouth into a rounded shape, one might expect the reward stimulus to have different reinforcing properties depending on the relatedness of the sound to the infant's behavior. We test this possibility in Experiment 5, choosing our reinforcing stimuli to be vowels that either share properties with the infants' behavior (/u/ sounds match the rounded lip shape) or do not share properties with the behavior (/i/ sounds mismatch the rounded lip shape).

3.1.3 Current experiment: Testing learning and memory for vowels using HAS

The design for Experiment 5 was modeled after a Near-Infrared Spectroscopy study (Benavides-Varela, Gomez, Macagno, Bion, Peretz & Mehler, 2011, described in detail in Chapter 4), in which newborn infants' memory for speech sounds was investigated using a familiarization-test procedure. Adapting the Benavides-Varela et al. (2011) protocol to a behavioral design, Experiment 5 implemented the HAS procedure to familiarize infants to one of two speech sounds, and later tested their memory for the previously heard speech sound: either an /u/ vowel (as in the word 'boot'), or an /i/ vowel (as in the word 'beet'). The study consisted of 3 phases: a 6-minute familiarization phase (/u/ or /i/), a 2-minute rest phase (which occurred in silence), and a 3-minute test phase⁵ (*same* sound for Control condition infants or *switch* sound for Experimental condition).

Because /u/ sounds are produced with rounded lips, which match the shape of an infant's mouth when sucking on a pacifier, it was hypothesized that being familiarized to /u/ would result in enhanced learning of the contingency. High Amplitude suck data (average per minute) and average suck amplitude were the two dependent measures for the following analyses, and were each analyzed in two ways: a) average first 2 familiarization minutes vs average last 2 familiarization minutes; and b) first test minute vs second test minute⁶. Analysis a) will determine if infants showed a pattern of sucking during familiarization that differed between the vowel to which infants were being familiarized (one that matched their lip shapes—/u/, compared to one that did not match their lip shapes—/i/); if infants learned the contingency between their suck behavior and the vowel-sound reward, there should be a greater number of sucks in the first two minutes compared to the last two minutes of familiarization. For analysis b), the data will be split by the following conditions: familiarization vowel (/u/ or /i/) and

⁵ Only the first two minutes of the three-minute test period were analyzed, due to the fact that infants' sucking behavior during the third test minute declined across all conditions.

⁶ As will be discussed in the General Discussion, we did not include the familiarization phase in the test-phase analysis, as this design did not implement habituation with a particular satiation criterion; therefore, we did not expect a systematic increase or recovery in sucking in comparison to the last familiarization minutes for infants in the experimental group (as in Eimas et al., 1971; Williams & Golenski, 1978; see Floccia et al., 1997)

condition (experimental or control), and will determine whether infants showed a memory response during the test phase that differed between the two kinds of vowels. Typically, infants in experimental conditions (who hear a different sound at test) exhibit greater sucking after the sound switch compared to infants in control conditions.

3.2 Experiment 5: Sensorimotor influences on contingency learning of vowel sounds in neonates

3.2.1 Method

3.2.1.1 Participants

The participants in Experiment 5 were 24 neonates (n=12 males, n=12 females) with normal hearing and no documented health problems. Infants had a mean age of 1.58 days (ranging from 0-4 days), and had an average 62.50 % English exposure while in utero as reported by parents (ranging from 0%-100%). In addition to the 24 infants included in the final analyses, data from 15 infants were not included due to falling asleep (n = 5), crying or fussiness (n = 2), equipment failure (n = 6), and parental interference (n = 2).

The infants' parents were recruited from the maternity wards at BC Women's Hospital in Vancouver, BC. Upon arrival to the testing room at BC Women's Hospital, the parents received an explanation of the study and procedure, and gave written consent for their infant's participation. After the study, the infants received a t-shirt as a thank-you for participating.

3.2.1.2 Stimuli

The vowel stimuli were a subset of the sounds used in Yeung & Werker, 2013 (originally provided by Rebecca Baier, Bill Idsardi, & Jeff Lidz), and included 5 unique tokens each of /i/ and /u/. Stimuli were recorded by a native female English speaker, and were normalized for

intensity. Because these two vowels are produced when the tongue is high in the mouth (and thus both have low first formants, F1), but differ mainly in front/backness of the tongue (/i/ is a front vowel and has a higher second formant, F2, than /u/, a back vowel with a lower F2), tokens were compared for F1 and F2. Acoustic analyses confirmed that /i/ and /u/ tokens differed primarily by F2, where /i/ had a higher F2 (M = 3022 Hz, SD = 40 Hz) than /u/ (M = 990 Hz, SD = 85 Hz). The average durations of the tokens of the two vowels were 454 ms for /i/, and 504 ms for /u/.

3.2.1.3 Apparatus

The high-amplitude suck apparatus included the use of a Phillips Soothie® pacifier (0-6 month old), a pressure transducer, and custom-built HASware (Molavi, Yeung, Byers-Heinlein, & Werker, in prep). Each Soothie pacifier was sterilized prior to use, and was attached to an adjustable microphone stand that contained silicone tubing. Each time the newborn sucked on the pacifier, this tubing carried the pressure-change information in the nipple to the pressure-transducer. The suck amplitude (psi) and timing of each suck were measured by the HASware. The infants were placed in a bathing chair during the study, in order to reduce head movements. Each infant experienced a one-minute baseline period in which she sucked in silence; this minute was used to determine the infant's high-amplitude suck threshold to be used during the study (the upper 80% of suck amplitude range).

3.2.1.4 Procedure

Newborns completed the study in a sound-attenuated testing room at BC Women's Hospital. The infants lay in a bathing chair placed within their own bassinets, and were situated approximately 5 feet (from speakers to ear) from the computer set-up and speakers. The experimenter placed the sterilized Soothie pacifier on the mechanical arm and introduced the

pacifier to the infant; infants were allowed to become familiar with the pacifier for a few moments (until they began to exhibit a consistent suck response). As mentioned, infants began with a one-minute baseline measure of their suck behavior in silence. The high-amplitude threshold was determined, and the study began. The experimenter did not touch or interact with the infant, except to immediately place the pacifier into the infants' mouths if it had been spit out.

During the familiarization phase, each time infants sucked at a level that exceeded their individual HA thresholds, they were presented with a vowel sound. The delay in the onset of sound presentation was .1 seconds (100 milliseconds) after the end of the high amplitude suck. In the event of a suck burst (when infants sucked more quickly than sounds could be presented), the sounds were presented in succession. The two-minute rest phase followed, during which time infants were presented with *no* sounds. Finally, during the test phase, infants were again presented sounds contingent on their high-amplitude suck behavior for a three-minute period.

Infants were randomly assigned to one of four conditions: Experimental-/u/ condition, Control-/u/ condition, Experimental-/i/ condition, or Control-/i/ condition. Each condition included the 6-minute familiarization phase during which infants' high-amplitude sucks were reinforced with a sound, the 2-minute rest phase during which infants' sucks were NOT reinforced with a sound (sucked in silence), and the 3-minute test phase during which highamplitude sucks were again reinforced with a sound. In the experimental-/u/ condition, infants were familiarized to /u/, and tested on /i/; in the control-/u/ condition, infants were familiarized to /u/, and tested on /u/; in the experimental-/i/ condition, infants were familiarized to /i/, and tested on /u/; in the control-/i/ condition, infants were familiarized to /i/, and tested on /i/.

3.2.2 Results

3.2.2.1 High amplitude sucks per minute

Familiarization phase: In order to test for learning effects between the two familiarization vowels, high-amplitude suck data for the 24 infants were averaged over the first 2 minutes of familiarization (minutes 1 and 2 were not significantly different from each other for both vowel types, F(1,22) = .53, p = .47), and the last 2 minutes of familiarization (minutes 5 and 6 were not significantly different from each other for both vowel types, F(1,22) = .41, p = .52). A 2 (Phase) X 2 (Familiarization Vowel) mixed ANOVA was performed on the HAS data for the within subjects factor of Phase (average first 2 familiarization minutes vs average last 2 familiarization minutes) and the between subjects factor of Familiarization Vowel (/u/ vs /i/). There was no main effect of Phase, F(1,22) = .23, p = .64, $\eta_p^2 = .010$, but there was a marginally significant interaction between Phase and Familiarization Vowel, F(1,22) = 3.64, p = .070, $\eta_p^2 = .14$.

Further inspection of means showed that infants who were familiarized to /u/ decreased their number of high amplitude sucks from the first 2 minutes (M = 48.29, SD = 15.90) to the last 2 minutes (M = 41.29, SD = 13.07) of familiarization compared to infants familiarized to /i/, who increased in the number of high amplitude sucks from the first 2 minutes (M = 48.38, SD = 15.46) to the last 2 minutes (M = 52.58, SD = 16.15) of familiarization. Infants familiarized to /i/ did not differ in number of high amplitude sucks in the first 2 minutes of familiarization compared to infants familiarized to /u/, t(22) = 0.013, p = .99; however, /i/-familiarized infants did exhibit marginally greater sucking during the last 2 minutes of familiarization compared to the infants familiarized to /u/, t(22) = 1.88, p = .069, 95%CI of the difference [-1.15, 23.73] (see Figure 3.1).



Figure 3.1. Familiarization phase HAS data split by familiarization-vowel. Error bars denote standard error of the mean.

Test phase: The two-minute test period was analyzed to investigate memory responses in the newborns, specifically whether HA suck behavior differed depending on the Condition (Experimental or Control) and Familiarization Vowel (/u/ or /i/). A 2 (Phase) X 2 (Condition) X 2 (Familiarization Vowel) mixed ANOVA was performed on HAS data using the within subjects factor of Phase (Test Minute 1 vs Test Minute 2) and the between subjects factors of Condition (Experimental vs Control) and Familiarization Vowel (/u/ vs /i/). There was a significant interaction between Phase and Condition, F(1,20) = 4.84, p = .040, $\eta_p^2 = .20$; there was no main effect of Phase, F(1,20) = .078, p = .78, $\eta_p^2 = .004$, and no interactions between Phase and Familiarization vowel, F(1,20) = .045, p = .83, $\eta_p^2 = .002$, nor between Phase, Condition, and Familiarization Vowel, F(1,20) = .066, p = .80, $\eta_p^2 = .003$. Follow-up analyses on the Phase by Condition interaction showed that while infants did not differ in the number of high amplitude sucks in the first test minute (control condition, M = 44.50, SD = 18.28; experimental condition,

M = 46.58, SD = 19.94), t(22) = .27, p = .79, infants in the experimental group had a significantly larger number of high amplitude sucks during the second test minute (M = 53.75, SD = 18.14) compared to infants in the control condition (M = 35.25, SD = 20.68), t(22) = 2.33, p = .029, 95%CI for the difference = [2.03, 34.97] (see Figure 3.2). These results suggest that during the two-minute test period, the infants in the experimental condition (regardless of the familiarization vowel) exhibited a greater number of sucks in response to the change in sound after the two-minute delay, while those infants in the control condition decreased their sucking behavior in the second test minute.



Figure 3.2. Test phase HAS data split by experimental condition.

3.2.2.2 Average suck amplitude

Because the HASware (Molavi et al., in prep) used in Experiment 5 also collected amplitude data for each suck, an exploratory analysis was conducted to investigate whether suck amplitude would be a valid dependent measure for detecting contingency learning and/or a memory response in newborn infants. The possibility exists that suck amplitude may offer a sensitive measure for detecting learning and memory, without categorizing sucks based on a predetermined threshold. While some researchers have used ratios of high amplitude sucks to all sucks as a dependent measure (Floccia et al., 1997), the current analysis takes into account the amplitude of all sucks within all minutes of the study (regardless of their reinforcing properties).

Familiarization phase: As in the HAS data analysis, the average suck amplitude for the infants in the first two minutes of familiarization was compared to the average suck amplitude for the last two minutes of familiarization to investigate whether suck amplitude is a sensitive dependent measure for learning effects. A 2 (Phase) X 2 (Familiarization Vowel) mixed ANOVA was performed on the suck amplitude data for the within subjects factor of Phase (average first 2 familiarization minutes vs average last 2 familiarization minutes) and the between subjects factor of Familiarization Vowel (/u/ vs /i/). There was no main effect of Phase, F(1,22) = .86, p = .36, $\eta_p^2 = .038$, nor was there was an interaction between Phase and Familiarization Vowel, F(1,22) = .013, p = .91, $\eta_p^2 = .001$. Regardless of the vowel to which infants were familiarized, suck amplitude in the first two minutes (M = .044 psi, SD = .012) was no different than the suck amplitude in the final two familiarization minutes (M = .042 psi, SD = .010). Thus, suck amplitude did not reveal any differences in suck behavior in the familiarization (learning) phase.

Test phase: The two-minute test period was analyzed to investigate memory responses in the newborns, specifically whether suck amplitude differed depending on the Condition (Experimental or Control) and Familiarization Vowel (/u/ or /i/). A 2 (Phase) X 2 (Condition) X 2 (Familiarization Vowel) mixed ANOVA was performed on suck amplitude data using the within subjects factor of Phase (Test Minute 1 vs Test Minute 2) and the between subjects factors of Condition (Experimental vs Control) and Familiarization Vowel (/u/ vs /i/). There was a significant interaction between Phase and Condition, F(1,20) = 4.55, p = .045, $\eta_p^2 = .19$; there

was also a main effect of Phase, F(1,20) = 5.17, p = .034, $\eta_p^2 = .21$. There was no interaction between Phase and Familiarization vowel, F(1,20) = .95, p = .34, $\eta_p^2 = .046$, nor between Phase, Condition, and Familiarization Vowel, F(1,20) = .10, p = .75, $\eta_p^2 = .005$. Follow-up analyses on the Phase by Condition interaction showed that while infants in the experimental condition did not differ in suck amplitude between the first test minute (M = .041 psi, SD = .010) and the second test minute (M = .041 psi, SD = .011), infants in the control condition significantly decreased their suck amplitude from the first test minute (M = .044 psi, SD = .0068) to the second test minute (M = .037 psi, SD = .0096), t(11) = 2.97, p = .013, 95%CI for the difference [.0018, .012]. These results suggest that infants in the experimental condition maintained a constant suck amplitude in response to the vowel change during the two-minute test period, while infants in the control condition decreased their suck amplitude across the two-minute test period when they heard the same vowel during test.

3.3 General discussion

Taken together, the results from the HAS analyses and the suck amplitude analyses suggest that newborn infants can discriminate and exhibit memory for the distinct vowel categories /u/ and /i/ after a brief period of familiarization. Evidence for the hypothesized sensorimotor effect on learning and memory in this behavioral task is limited; the only marginal difference related to the vowel-articulatory match or mis-match occurred during the familiarization phase HAS analyses, in which infants who were familiarized to /u/ had a lower HAS rate in the last two familiarization minutes compared to infants familiarized to /i/. However, given the test phase results, even if newborns are processing the /u/ and /i/ sounds differently while learning the contingency between their suck behavior and a reinforcing sound, they show the same patterns of memory response in the test minutes.

Regardless of the vowel to which they were familiarized, newborns who were presented a different vowel during the test minutes showed a larger number of high amplitude sucks in the second minute than those who heard the same vowel during test. This suggests that newborns can discriminate these two vowels, and that they can remember them even after a two-minute delay period. These findings support previous reports showing memory responses in newborns and preverbal infants in the first few months of life (Benavides-Varela et al., 2011; Haley, Weinberg, & Grunau, 2006; Jusczyk, Pisoni, & Mullennix, 1992; Swain et al., 1993). The fact that infants can discriminate and show memory for single vowel sounds also adds to the large literature on speech perception capabilities in newborn infants (eg, Bertoncini et al., 1988; Bertoncini, Floccia, Nazzi, & Mehler, 1995; Dehaene-Lambertz & Pena, 2001; McAdams & Bertoncini, 1997; Moon et al., 2013; Pena et al., 2003).

3.3.1 Memory for and discrimination of vowels

A discussion of the test analysis is required, given our decision to exclude the suck behavior from the previous phases of the experiment in investigating the memory response (in contrast to Bertoncini et al., 1995; Eimas et al., 1971; Floccia et al., 1997; Floccia, Nazzi, & Bertoncini, 2000; Jusczyk & Derrah, 1987; Jusczyk et al., 1992; Mehler et al., 1988). While previous reports using HAS have suggested that the trained contingency between the suck behavior and vowel reinforcer is essential to show discrimination in an *increase* in sucking from the pre-shift minutes to the post-shift minutes (Williams & Golenski, 1978; Floccia et al., 1997), we did not implement a design that required infants to reach a satiation criterion. As has been suggested elsewhere (Floccia et al., 1997), it is difficult to see increases in suck behavior even above the baseline minutes of sucking during which the pacifier is novel; to do so, infants typically must experience a change in the sound stimulus at "an unusually low point of sucking activity" to show an increase in their suck behavior in a speech task (Floccia et al., 1997, page 195). Given that neither /u/-familiarized nor /i/-familiarized infants significantly decreased their suck behavior during familiarization, we chose to analyze the test phase separately.

In the test phase data, we showed a *decrease* in sucking across the test minutes for infants in a control condition. Cowan and colleagues (Cowan, Suomi, & Morse, 1982) provide similar evidence using a modified version of the HAS procedure. They implemented a design in which one group of infants experienced alternated presentation of two sound stimuli compared to a group of infants who heard the same sound throughout a fixed length of familiarization, and found that infants who experienced alternating minutes maintained their high-amplitude suck rate, while infants who were exposed to the same sound decreased their sucking rate (Cowan et al., 1982; Vouloumanos & Werker, 2007b). In the present data, although we did not implement the same alternating procedure, we did find that compared to infants who experienced a shift in the reinforcing stimulus, infants who heard the same vowel at test decreased their suck rate across the test minutes. Even without requiring a satiation criterion in the first phase of the experiment, we showed evidence that infants discriminated the vowels /u/ and /i/.

3.3.2 Suck amplitude as a dependent measure

Because the HASware used in Experiment 5 allowed us to collect amplitude data for the suck behavior across the 11 minute study, we performed an exploratory analysis to determine whether suck amplitude is sensitive to either the contingency learning (familiarization) or the memory/discrimination (test) phases of the experiment. We found that in the test phase of the experiment, suck amplitude data corroborated the HAS findings: infants who were in a control condition (regardless of the vowel to which they were familiarized) decreased their average sucking amplitude during the two-minute test period. In contrast, those in the experimental

condition maintained a constant suck amplitude between the two test minutes. These findings suggest that suck amplitude data may be an additional dependent measure to consider in high-amplitude sucking experiments that investigate memory and discrimination after a period of familiarization. Based on the data presented here, infants who continued to hear the same sound after a short delay decreased their suck amplitude, exhibiting declining interest in the task.

We do not suggest that a HAS analysis and an analysis of suck amplitude be presented together; suck amplitude is a condition upon which the high-amplitude suck threshold is based, so these two measures are not independent of one another. We include both in this report to introduce suck amplitude as a potentially useful dependent measure in lieu of HAS counts; it accounts for the total range of suck activity by the infants, rather than being based on a predetermined criterion for what is considered a 'proper' suck. In studies of the non-nutritive suck behavior of premature infants, suck amplitude is a measure used to quantify oromotor development and function (Stumm, Barlow, Vantipalli, Finan, Estep, Seibel, Urish, Fees, Poore, Cannon, & Carlson, 2005). We suggest that it may be a useful measure to quantify cognitive functioning and perception in neonates. In much the same way that pupillometry offers an additional dependent measure for attention in cognitive tasks (including in looking time in infants) (Laeng, Sirois, & Gredeback, 2012), we reasoned that average suck amplitude may be able to quantify a novelty response with a larger range of possible responses than is possible by counting the total number of high amplitude sucks. Further research confirming these patterns is clearly necessary.

3.3.3 Sensorimotor effects and 'contingency awareness'

The only indication of a sensorimotor effect in this contingency learning – speech discrimination study was found in the marginal vowel-effect in the familiarization phase. This

marginal interaction suggests that infants familiarized to /u/ may be exhibiting different patterns of sucking compared to infants familiarized to /i/; although only marginally significant, /u/- familiarized infants tended to decrease their high amplitude suck behavior as they continued to receive the vowel /u/ as a reinforcer, while /i/-familiarized infants tended to increase their suck behavior in response to the /i/-reinforcement.

Further, the sucking patterns exhibited by both /u/-familiarized and /i/-familiarized infants seem to be in line with previous contingency-learning research. For example, some studies (mainly those with fixed length experimental periods, as in Siqueland & DeLucia, 1969; Trehub & Chang, 1977) show that the contingent presentation of stimuli that reinforce high-amplitude suck behavior results in an *increase* in sucking across a period of time—usually no longer than a 6-minute period. Others, typically those that implement habituation designs, show that after an initial increase in sucking in the first post-baseline minutes, a *decrease* in high-amplitude sucks occurs with contingent presentation of speech sounds—which can take upwards of 9 to 10 minutes—until infants reach a predetermined satiation point (Eimas et al., 1971; Floccia et al., 2000; Jusczyk et al., 1988). Therefore, while some evidence suggests that successful learning is portrayed by an eventual decrease in sucking, others purport that learning is exhibited by an increase in suck behavior.

One must keep in mind the fact that Experiment 5 implemented a familiarization—not a habituation—design. This choice in experimental procedure determined that infants would experience the same length of study, but did not control for stimuli presentation across infants, nor did it require infants to decrease their sucking rate by a certain criterion as in other HAS designs (Eimas et al., 1971; Williams & Golenski, 1978). However, in the present data, it may be the case that infants who are familiarized to /u/ (and decrease in their suck behavior) have

reached satiation more quickly than /i/-familiarized infants (who increase in their suck behavior), providing some evidence that contingency learning (ala Watson's idea of 'contingency awareness', Watson, 1966) unfolds differently when the sound reinforcers share properties with the reinforced behavior.

3.3.4 Conclusions

Experiment 5 provided strong evidence for newborns' memory and discrimination capabilities for short vowel sounds, and suggestive evidence that these capabilities are mediated by sensorimotor information provided by the shape of the articulators. While these findings certainly add to the research on speech perception and discrimination in neonates, particularly in their ability to distinguish and remember speech stimuli of such short duration, additional research is necessary to determine whether suck behavior—and a rounded mouth shape reliably contributes to speech processing in neonates. In Chapter 4, I further investigate this question using neuroimaging technology, to identify whether the sensorimotor link is evident in underlying neural activity during the high-amplitude suck procedure.

Chapter 4 : Neural networks involved in processing vowel sounds in neonates 4.1 Introduction

4.1.1 Neuroimaging techniques in infancy

Behavioral techniques, including the high amplitude suck procedure used in Chapter 3, have helped researchers discover the processing capabilities of newborn infants; in recent years, the use of neuroimaging techniques has expanded our understanding of cognitive processing in the newborn, as well as early patterns of brain activation (Aslin, Shukla, & Emberson, in press; Gervain, Mehler, Werker, Nelson, Csibra, Lloyd-Fox, Shukla, & Aslin, 2011; Kuhl, 2010). Very generally, neuroimaging techniques either directly detect the electrical activation in the brain in response to a stimulus, or measure a hemodynamic response (HDR) that occurs as a consequence of neural activity. Concerning the area of language acquisition, developmental neuroscientists employ Electroencephalography/Event-Related Potentials, or EEG/ERP (e.g. Conboy, Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2008), functional Magnetic Resonance Imaging, or fMRI (e.g. Dehaene-Lambertz et al., 2002 & 2006), Magnetoencephalography, or MEG (e.g. Imada et al., 2006), and Near Infrared Spectroscopy, or NIRS (Lloyd-Fox, Blasi, & Elwell, 2010; Gervain et al., 2011) in the first year of life to study the linguistic processing capabilities of preverbal infants.

Each neuroimaging technique mentioned above has a set of advantages and disadvantages for use with developmental populations, depending on cost, sensitivity to movement, and temporal and spatial resolution in response to stimuli (Gervain et al., 2011; Kuhl, 2010; Lloyd-Fox et al., 2010). The current set of experiments utilize NIRS, which is less susceptible to corruption by movement artifacts and has a higher spatial resolution than EEG; it has high temporal resolution and is silent, making it more useful for auditory presentation

compared to fMRI; and is relatively inexpensive compared to MEG systems. However, NIRS has lower temporal resolution than EEG, and lower spatial resolution than fMRI, and does not provide or record anatomical data (of the underlying brain structures from which neural activation is being recorded) as can be measured using MRI. Yet, for certain questions, especially those that pertain to language processing in newborn and preverbal infants, NIRS has proven to be a useful tool for studying the neural activation exhibited by infants across a range of stimuli types, areas of interest, and age groups (see Lloyd-Fox et al., 2010, Gervain et al., 2011, and Aslin et al., in press, for reviews on the published works using NIRS with infants).

NIRS systems use near-infrared (NIR) light to detect changes in concentration of oxygenated and deoxygentated hemoglobin (oxyHb, deoxyHb, respectively) in the blood (Jobsis, 1977); the NIR light sources are coupled with light detectors to create channels at different regions in which the HDR can be determined. When near-infrared light sources are placed on the skull of a newborn infant, changes in blood-oxygenation at the level of the cortex can be measured as an index of neural activity; because the oxyHb and deoxyHb chromophores have different absorption properties of near-infrared light, blood oxygenation can be measured via the attenuation of NIR light from the source to the detector in each channel. Due to the thickness of newborns' surface tissues, NIR light penetrates up to 30mm deep into the head, reaching about 10-15mm into the newborn cortex in a 'banana' shaped trajectory. As mentioned, the various tissues and medium through which the NIR light passes each have different absorption qualities (including the oxyHb and deoxyHb chromophores), as some tissues absorb more light than others; these absorption characteristics can be accounted for in optical imaging techniques (as in NIRS) using a modified version of the following Beer-Lambert law:

$$A = -\log (I/Io) = (c \ge \lambda \ge l)$$

where A is the absorbance of the light, I is the intensity of the transmitted NIR light *after* passing through the tissues, Io is the intensity of the NIR light *before* it passes through the tissues, c is the density of the tissues, $\varepsilon \lambda$ is the 'molar extinction coefficient characteristic' of the tissues for a light with a wavelength λ , and l is the distance that the light travels through the tissues. Because the initial intensities of the NIR light are known, as are the qualities and extinction coefficient characteristics of the oxyHb and deoxyHb chromophores through which the light passes, the exiting intensity of the NIR light can be used to calculate *changes* in concentration of the oxygenated hemoglobin; this requires a modified version of the Beer-Lambert Law which also accounts for the 'scatter' that occurs as NIR light crosses tissues such as skin and bone. These computations result in measures of the concentration changes of oxyHb and deoxyHb, which can then be used as indicators of (the related) underlying neural activity.

A typical HDR as measured by NIRS involves an increase in concentration of oxyHb after the onset of a stimulus (and neuronal activation), coupled with a lesser decrease in deoxyHb. While the BOLD signal in fMRI more closely relates to the deoxyHb signal in NIRS in adult studies (Huppert, Hoge, Diamond, Franceschini, & Boas, 2006), deoxyHb results are less consistent in infant studies; as a result, oxyHb is the measure most commonly reported in infant research using NIRS (Lloyd-Fox et al., 2010).

Given the temporal characteristics of a typical HDR—a metabolic, slow correlate of brain activation—most NIRS studies implement a blocked design. In blocked designs, stimuli are presented to the infant for a period of up to 30 seconds (in the stimulation blocks) during which the HDR is expected to occur. The stimulation block is then followed by a control block to allow the HDR to return to a baseline level; the control block typically involves no (or minimal)

stimulation. Typical response latencies post-stimulus onset are several seconds in length, and after the response plateaus, the return to baseline takes an additional 5 to 10 seconds. With a blocked design, the slow HDR can be accounted for when building block lengths, and the HDR during the stimulation blocks can be compared to the response in the control blocks. However, implementing event-related designs using NIRS would allow researchers to investigate responses to relatively shorter periods of stimulation; event-related NIRS designs may help identify neural areas involved in speech sound processing independent of task or attention-related effects (Aslin et al., in press; Gervain et al., 2011).

While many studies involving NIRS with newborn infants have utilized block designs in which activation patterns are recorded during alternating blocks of different stimuli types, a recent NIRS study has adapted a familiarization-test blocked design (Benavides-Varela et al., 2011). In this study, newborns were familiarized to a single nonsense word that was repeated 6 times in a single stimulation block (which lasted 10 seconds on average). The familiarization phase included a series of 10 blocks of stimulation (with 25-35 second silent blocks in between stimulation) for a total of 6 minutes. After familiarization, infants experienced a 'rest phase' of silence for 2 minutes. During the test period, infants were presented 5 blocks of either the familiar word or a novel word. Results showed that, within the first block of the test phase, newborns who were presented the novel word showed a greater change in concentration of oxyHb compared to infants who were presented the familiar word; the researchers interpreted the increase in oxyHb to the novel nonsense words as a memory response. These differences in the HDR were found bilaterally in temporal-parietal and anterior areas. This familiarization-test NIRS paradigm offers a way for researchers to investigate discrimination, preference, learning, and memory processes in newborns, as well as the general neural areas involved.

4.1.2 Neural evidence for sensorimotor influences on speech processing in infancy

Turning now to the specific topic of interest, sensorimotor influences on speech perception in infancy, recent evidence suggests that pathways involving the motor system are active in preverbal infants. Dehaene-Lambertz and colleagues (Dehaene-Lambertz et al., 2002) identified brain regions involved in language processing in 3-month-old infants using fMRI: relative to silent periods, the presentation of speech activated the left temporal lobe, bilateral superior temporal sulci, and the left planum temporale. Follow-up work indicated that sentences presented to 3-month-old infants activated a network of perisylvian areas, including the superior temporal regions, as well as inferior frontal regions; in adults, the left inferior frontal gyrus (Broca's area) has been implicated in the production of speech and inner speech, making the finding of activity in this area in preverbal infants somewhat surprising (Dehaene-Lambertz et al., 2006). Using MEG, Imada and colleagues (2006) have also found evidence of Broca's area activation during the perception of speech in 6-month-old infants. These findings suggest that speech perception activates an area involved in the production of speech in adults, and that an early perceptuo-motor link exists in infants for the processing of speech information. The authors argue that this link requires experience to bind the speech (perception) and motor (action) areas, as they did not find inferior frontal region activation in newborns (Imada et al., 2006). In fact, recent evidence concerning the effect of native language experience on neural patterns has shown that auditory speech information can differentially activate auditory and motor neural areas between 7 and 12 months of age: while 7-month-olds exhibit similar patterns of activation in auditory and motor areas to both native and non-native speech, in 12-month-olds, non-native sounds activate motor areas to a greater degree than native speech sounds, and native sounds result in greater activation in auditory areas compared to non-native sounds (Kuhl et al., 2014).

Using Diffusion Tensor Imaging (DTI), other researchers have found that two functionally distinct networks exist for language processing at birth: a ventral pathway that connects the inferior frontal gyrus with the superior temporal cortex (and is responsible for speech comprehension), and a dorsal pathway that connects the temporal cortex to the premotor cortex (Perani et al., 2011, see Figure 4.1). Research with adults has shown similar pathways, both ventral and dorsal, for language processing (Hickok & Poeppel, 2004, 2007; Saur et al., 2008). The dorsal pathway that exists in newborn infants, the authors argue, allows auditory-tomotor mapping to occur, particularly during speech perception. A second dorsal pathway, which is not fully myelinated until later in life, more directly connects the temporal cortex to Broca's area; this dorsal pathway supports the processing of more complex linguistic stimuli (such as sentences with complex syntax) that is more relevant later in development (Friederici, 2012). In line with previous arguments, these researchers also state that the dorsal pathway evident in newborns "guarantees sensory-to-motor feedback during the infant's babbling phase during the first months of life" (Perani et al., 2011, page 16060).

Figure 4.1. Ventral and dorsal language pathways in adults and neonates.

Fiber tracking of DTI data for (A) adults and (B) newborns. Two dorsal pathways are present in adults—one connecting the temporal cortex to Broca's area (blue), and one connecting the temporal to the premotor cortex (yellow). In newborns, only the dorsal pathway to the precentral gyrus can be detected. The ventral pathway connecting the ventral inferior frontal gyrus to the temporal cortex (green) is present in adults and newborns (From Perani et al., 2011).



However, while these studies suggest that the necessary networks for sensorimotor mapping are in place for processing language at birth, none of these experiments directly investigated such a sensorimotor link; the kind of experience necessary to link perception and action during infant speech perception remains an open question. As discussed in previous chapters, there is evidence to suggest that articulatory-motor gestures, including a sucking behavior that forces the lips into a rounded shape, versus using a teething toy to spread the lips, influences the perception of audiovisual speech in 4-month-old infants (Yeung & Werker, 2013). Taking these findings into account, experience with speech production per se may not be necessary to form this link; instead, it may be that the sensorimotor influence on speech perception exists much earlier in development, but that relevant, matching information from the articulators may be required to detect this influence. Assuming that the infants' articulatory systems (i.e. mouth shapes) were not manipulated in the neuroimaging studies described in the previous paragraph, a more convincing argument for such a perceptuo-motor link would come from studies in which the movements of the mouth were controlled while listening to speech—if an early link exists, one should expect differential activation in dorsal processing stream areas in cases where an infant's articulators either provide relevant, matching information about the speech she is hearing, compared to cases where the articulators provide irrelevant, mismatching information about the speech she is hearing.

Building on those findings presented in Experiment 5 (Chapter 3) in which a sensorimotor influence was not reliably evident in the behavioral results in a speech perception (memory) task, the experiments reported in this chapter concern the neural underpinnings of the HAS task, and identify whether certain speech sounds are processed differently given the shape of the infants' mouths during the procedure. Therefore, taking into account the findings from Experiment 5 (Chapter 3), the many studies utilizing HAS as a behavioral technique in newborns, and the recent advances in neuroimaging techniques used in developmental research, Experiments 6 and 7 attempted to co-register the use of NIRS with HAS, such that the presentation of sounds to each infant was either contingent (Experiment 6), or not contingent, (Experiment 7) on his/her own suck behavior. In doing so, this created an inherently "event-related" design: the presentation of speech sounds (both in timing and number of stimuli across the study) in a HAS paradigm is different for each infant. Thus, the ideal analysis of HAS-related NIRS data would be one in which HDRs could be deconvolved into separate signals for

each sound stimulus. However, few studies have used event-related designs in infants (but see Jasdzewski, Strangman, Wagner, Kwong, Poldrack, & Boas, 2004; and Chen, Vaid, Bortfeld, & Boas, 2008, for event-related designs in adults using NIRS); to our knowledge, no studies have implemented a design in which every infant experiences a different stream of sound presentation. To account for these difficulties in analyzing HAS-presented NIRS data, we used a relatively new multivariate analysis technique—Constrained Principal Component Analysis—to analyze the NIRS data.

4.1.3 Constrained Principal Component Analysis: Analyzing event-related designs

To determine the neural activation in the newborns in Experiments 6 and 7, we utilized Constrained Principal Component Analysis (CPCA), a multivariate analysis method that is able to take into consideration the variable stimulus presentations within and across individual infants, as well as the infants' HDRs. CPCA treats the stimulus presentations as predictor variables, and the HDRs as criterion variables, and combines this regression analysis with principal component analysis into an integrated framework (Hunter & Takane, 2002; Metzak, Feredoes, Takane, Wang, Weinstein, Cairo, Ngan, & Woodward, 2011; Takane & Hunter, 2001). Very generally, CPCA involves two analyses. The first, the 'external' analysis, partitions the total variability that exists in the HDR data into a proportion of variability that can be accounted for by the stimulus presentations (the predictor variables), and the variability that is independent of them (error). This external analysis can be summarized in the equation:

Z = GC + E

The Z matrix is a matrix that contains HDR data for all subjects for all channels; the rows in Z correspond to the individual time points of data collection for all subjects stacked on top of one another, and the columns in Z represent the number of channels (24); therefore, the first row of

the Z matrix corresponds to the first time point for the first subject, and the numbers in each of the columns correspond to the oxyHb concentration change value for each of the 24 channels. The G matrix is the 'design' matrix, which takes into account the stimulus (sound) presentation for all time points for each infant, and provides a model of the expected HDRs for all time points. The G matrix is the same length as the Z matrix (accounting for all time points for all subjects), while the width of G (in the case of the current data) is determined by the *expected* length of the hemodynamic response, and represents the peri-stimulus time; because previous studies typically report HDRs around 30 seconds, the width of G is 300 units (1 second per 10 units in G). A value of 1 is placed in the row and column of G where the NIRS HDR is to be estimated for the corresponding time point in Z, and 0s were placed in all other columns. At the beginning of a sound stimulus, that corresponds to row 400 in Z for example, row 400 & column 1 of G would contain a value of 1; all other columns in row 400 of G would contain 0s. In row 401 of G, only column 2 would contain a value of 1; all other columns would contain 0s. In this way, one can build a model that dictates in a binary manner whether and when activation in the Z matrix is expected in 30-second time periods.

The Z matrix is regressed onto the G matrix, and the output of this 'external' analysis is a set of stimulus-presentation-specific regression weights in the C matrix, as well as a matrix E, the error matrix that contains the variability in Z that is independent of G. Thus, in regressing Z onto G, this allows the (likely overlapping) HDRs in Z to be deconvolved into separate signals resulting from the sound stimulus presentations that vary between infants.

The second analysis, the 'internal analysis', extracts components that designate networks of functionally interconnected channel activations that relate to the stimulus presentations depicted in G. This internal analysis applies PCA to the GC matrix (which contains the

variability in the HDR that was predictable from the design matrix G) in order to determine underlying structures that are related to the pattern of activation across the 24 NIRS channels. The internal analysis utilizes singular value decomposition (SVD) as portrayed in the equation:

$$UDV' = GC$$

where U is the decomposed version of the Z matrix (left singular vectors, length of G x length of G), D is the diagonal matrix of singular values, and V contains the variance contributed by the 24 channels (right singular vectors, 24 channels x 24 channels). The V matrix contains information about the neural regions involved in the extracted components, and the percentage of variance accounted for by the extracted networks; the matrix VD contains channel-based component loadings. Finally, to determine the importance of the peri-stimulus time points in each column of G to the components extracted in the SVD equation above (across all channels), predictor weights are calculated as the matrix P in the equation

$$U = GP$$

by correlating the G matrix with the (rotated) U matrix⁷. These predictor weights can be plotted across time, and should resemble a HDR curve; further, they can be used to identify differences in activation in the separate networks for different experimental conditions.

4.1.4 Current experiments: Co-registration of HAS and NIRS

Building from the findings in Experiment 5 (Chapter 3), Experiments 6 and 7 concerned the neural areas involved in processing a) sounds whose presentation was contingent on the infants' suck behavior (Experiment 6), or b) sounds whose presentation was not contingent on infants' suck behavior, but were rather presented randomly (Experiment 7). Importantly, the sounds to which infants were exposed either matched the lip shape of the infants in the study (/u/

⁷ The GC matrix (and the U matrix) underwent Varimax rotation, *before* computing predictor weights in U = GP, which redistributes the variance in the GC matrix.

sounds), or mismatched the lip shape of the newborns (/i/ sounds). In these experiments, we aimed to identify networks involved in processing the /u/ and /i/ sounds during the learning/encoding phase (familiarization phase), and during the memory probing phase (test phase). Because the output of CPCA includes predictor weights for the extracted components, any differences found in the processing of these vowel sounds (during familiarization and test) would be represented in the levels of activation in the components between the experimental conditions. Given the findings mentioned earlier in this chapter (and in Chapter 1), we hypothesized that the extracted components would represent networks that involved the frontal and temporal-parietal regions, and that these networks would be sensitive to the vowel to which the infant was being exposed – resulting in different levels of activation during /u/-presentation compared to /i/-presentation. Thus, we hypothesized a significantly greater HDR in the auditory and motor areas (specifically in the motor cortex, near the frontal and temporal-parietal junction) during /u/-familiarization compared to /i/-familiarization, because of the motor-match in articulation and perceived sound.

4.2. Experiment 6: Neural networks involved during speech sound presentation *contingent* on suck behavior

The behavioral task and design for Experiment 6 were identical to the procedure used in Experiment 5 (Chapter 3). In Experiment 6, infants were presented either /u/ or /i/ vowels for a 6-minute familiarization phase; they then experienced silence for a 2-minute rest phase; finally, they were presented either a different vowel (experimental condition) or the same vowel (control condition) during the 3-minute test period. All sounds were presented contingently after a high-amplitude suck. Concurrently with the HAS procedure, NIRS measured the infants' neural response to the sounds.

4.2.1 Method

4.2.1.1 Participants

The participants in Experiment 6 were 16 neonates (n=8 males, n=8 females) with normal hearing and no documented health problems. Infants had a mean age of 1.13 days (ranging from 0-4 days), and had an average of 78.75 % English exposure while in utero as reported by parents (ranging from 0%-100%). In addition to the 16 infants included in the final analyses, data from 17 infants were not included due to falling asleep (n = 7), crying or fussiness (n = 6), equipment failure (n = 2), and failure to obtain sufficient analyzable NIRS data (n = 2).

The infants' parents were recruited from the maternity wards at BC Women's Hospital in Vancouver, BC. Upon arrival to the testing room at BC Women's Hospital, the parents received an explanation of the study and procedure, and gave written consent for their infant's participation. After the study, the infants received a t-shirt as a thank-you for participating.

4.2.1.2 Stimuli & apparatus

Speech sound stimuli (/u/ and /i/ vowels) were identical to those used in Experiment 5 (Chapter 3).

The NIRS system used in Experiment 6 was a Hitachi ETG-4000, with infrared light wavelengths of 690nm and 830nm, and source-detector separation of 3cm; by using two wavelengths of NIR light, this allowed a calculation of the relative change in light absorption, and thus, the relative concentrations of oxyHb and deoxyHb in the cortical tissue. Sampling rate of the NIRS signal recording was 10 Hz, and the Hitachi machine used a laser power of 0.75 mW. The 9 NIRS optical fibers (1 mm) were encased in two chevron-shaped silicone probes, consisting of 5 emitters (sources) and 4 detectors per probe. The source-detector configuration follows Gervain and colleagues' configuration (Gervain, Macagno, Cogoi, Pena, & Mehler,

2008), which had 12 channels per hemisphere (channels 1-12 on the left, channels 13-24 on the right) (see Figure 4.2a). The two chevron-shaped probes were placed on the sides of the newborn's skull and generally covered frontal, temporal, and parietal regions, and were held into place using a soft woven cap (see Figure 4.2b). NIRS data were collected throughout the study, and were marked for both the HA sucks, as well as the presentation of the sound stimulus, by signals sent from the HASware.

Figure 4.2. Representation of NIRS probe placement on a neonate skull.

a) Silicone probes (black fingers): red circles indicate sources, blue circles indicate detectors; green (numbered) squares indicate channels in which relative hemoglobin concentration was measured. b) Newborn participating in the study.



The high-amplitude suck apparatus was identical to the one used in Experiment 5 (Chapter 3), and involved the use of either a Phillips Soothie® silicone pacifier (0-6 month old)
or a Gerber First Essentials® Soft Center latex pacifier, a pressure transducer, and custom-built HASware (Molavi et al., in prep). Each pacifier was sterilized prior to use. After being capped with NIRS, the newborn was placed in a bathing chair in order to reduce head movements. Each infant experienced a one-minute baseline period in which she sucked in silence; this minute was used to determine the infant's high-amplitude suck threshold to be used during the study (the upper 80% of suck amplitude range).

4.2.1.3 Procedure

Newborns completed the study in a sound-attenuated testing room at BC Women's Hospital. Once in a state of quiet alertness, the infants were capped with the NIRS probes, covering frontal, temporal and parietal regions; although head size and circumference was variable between infants, probes were placed so that they reached each side of the forehead, and covered the area above the ears (up to a few centimeters above the top of the ear) (see Figure 4.2b). Newborns were then laid in a bathing chair placed within their own bassinets. As soon as the infant was comfortable in the bathing chair, the NIRS probes were calibrated, during which time proper placement and contact with the skull was confirmed; after calibration, the Hitachi ETG-4000 began collecting neural activation data. Infants were introduced to the pacifier, and as soon as they began to exhibit a consistent suck response, they experienced the one-minute baseline measure of their suck behavior in silence. The high-amplitude threshold was determined, and the study began. The experimenter did not touch or interact with the infant, except to immediately place the pacifier into the infants' mouths if it had been spit out.

During the 6-minute familiarization phase, each time infants sucked at a level that exceeded their individual HA thresholds, they were presented with a vowel sound at a level of 70 dB. The HASware recorded both the timing and amplitude of the suck as well as the timing of

presentation of the sound, and sent this information as markers to the Hitachi machine to provide a time-locked record of stimulus presentation and suck behavior in the NIRS data. The delay in the onset of sound presentation was .1 seconds (100 milliseconds) after the end of the high amplitude suck. During a suck burst (when infants sucked more quickly than sounds could be presented), the sounds were presented in succession. The two-minute rest phase followed, during which time infants were presented *no* sounds. Finally, during the test phase, infants were again presented sounds contingent on their high-amplitude suck behavior for a three-minute period.

Infants were randomly assigned to one of four conditions: Experimental-/u/ condition, Control-/u/ condition, Experimental-/i/ condition, or Control-/i/ condition. In the experimental-/u/ condition, infants were familiarized to /u/, and tested on /i/; in the control-/u/ condition, infants were familiarized to /u/, and tested on /u/; in the experimental-/i/ condition, infants were familiarized to /u/, and tested on /u/; in the control-/i/ condition, infants were familiarized to /i/, and tested on /u/; in the control-/i/ condition, infants were familiarized to /i/, and tested on /i/.

4.2.1.4 Data analysis and preparation of matrices

NIRS data from the Hitachi ETG-4000 (light absorption signals obtained by optical topography) were used to compute the change in hemoglobin (both oxyHb and deoxyHb) concentration signals for each of the 24 channels for each subject. Only relative oxyHb concentrations were used in the following analyses.

As described previously, NIRS data were analyzed using Constrained Principal Component Analysis. Given the procedural design of Experiment 6, each infant's NIRS data were first coded for the onset of each suck burst (and therefore, the onset of a series of speech sound presentations) for the three phases of the experiment⁸. Suck bursts were defined as periods of continuous sucking in which the individual sucks were separated by less than 2 seconds; a suck burst ended when a suck occurred and was followed by more than 2 seconds without another suck (DeCasper & Carstens, 1981). These suck-burst onset time values were used to build the design G matrix; neural activation (as a HDR) was expected during suck/sound presentation, and the model matrix assumed the hemodynamic response would be on the order of 30 seconds in length. Although CPCA would allow a G matrix to model responses to *each* suck/sound presentation, we chose to analyze based on suck bursts because the sounds were presented in such close succession.

Separate CPCAs were performed for each phase of the experiment. Given the different kinds of processing expected in each phase (learning in familiarization phase, and memory in the test phase), we chose to analyze these phases separately to identify the (potentially different) networks involved. Construction of the Z and G matrices were similar for each phase.

NIRS data from all subjects were used to construct the Z matrix. Rows of NIRS data in Z correspond to time (sampling rate of NIRS was 10 Hz; 1 row is 100 ms), and columns correspond to activation channel (1-24); values in each row designate oxyHb concentration during that time point within each of the 24 channels. Each subject's NIRS data matrix varied in length (because recording NIRS data began as soon as the infant was comfortable, there were varying lengths of time before the study began, thus variable number of rows for each subject's data), but each was 24 columns in width, corresponding to the 24 NIRS channels. Matrices were stacked one on top of the other, to create one Z matrix (time x channel).

⁸ Because of the close time-lock between suck and sound presentation, timing of sound presentation was assumed to similar enough to the suck time onset that data in Experiment 6 was analyzed coded only for suck onset (but see Experiment 7 below).

The design G matrix was the same length as the Z matrix, and as mentioned earlier, the width was determined by the expected time of the HDR—in this analysis, we used a 30 second HDR model. Individual G matrices were created for each subject, and were 300 columns wide (300 columns per 100 ms units = 30 seconds). Based on the time points of suck/sound presentation coded previously, a row of an infant's G matrix contained a value of 1 when the HDR was to be estimated (corresponding to the suck/sound response) and 0s in all other rows in a diagonal (the beginning of estimation starts in column 1 of row X in which a value of 1 is placed, and so on until the end of the 30 second period in column 300 of row X+299). This created a binary model of when a HDR was expected based on an individual infant's suck behavior/pattern of sound presentation. Each individual G matrix was placed into the larger G matrix (time x HDR response time). Z and G were standardized to have zero mean and unit standard deviation in each column for each subject.

The external (Z=GC+E) and internal analyses (UDV'=GC, U=GP) (described in Section 4.1.3 above) were conducted using a custom MatLab script for each experimental phase. Outputs of these analyses provided two separate components (and their component-loadings for each channel) that represented the neural networks involved in processing the suck/speech sound. The number of components was determined by using scree plots, in which the variance accounted for by each component was calculated and depicted as an eigenvalue on the Y axis, with all possible extracted components on the X axis (24 in this data set); the point at which the slope of the curve levels off indicates the number of components that should be generated by the analysis. Based on initial analysis of scree plots during each phase (in both Experiments 6 and 7), two components per phase were extracted. In addition, the predictor weights for each component (for each subject) were calculated to represent the HDR, and were used in the ANOVAs reported below. Finally, the amount of variance accounted for by the model in GC, and the variance accounted for by the components, were computed.

4.2.2 Results

4.2.2.1 Constrained Principal Component Analysis: Familiarization phase

Using CPCA, two components were extracted from the NIRS data collected during the familiarization phase. The percentage of variance in the NIRS data that was accounted for by the GC matrix was 6.92, and the percentages of variance in GC that were accounted for by components 1 and 2 were 6.02 and 1.88, respectively. The underlying areas of activation that correspond to each component are plotted in Figures 4.3 and 4.4; component-loadings were negative for each extracted network⁹. As described above, the average predictor weights (G loadings by subject) were plotted as a function of expected HDR time (30 seconds), and represent the response of the corresponding networks during the familiarization phase—when the infants learned the contingency between their suck behavior and sound reward—for the two different vowel sounds.

Component 1: Given the pattern of activation as plotted in Figure 4.3a, component 1

⁹ I discuss the implications of negative component loadings in the General Discussion below. In adult fMRI, negative loadings represent *deactivation* in network areas, and thus positive peaks in the HDR plots (from the predictor weights) would indicate *decreasing* concentrations of oxyhemoglobin. However, given the disagreement in the literature concerning the meaning of decreasing levels of oxyhemoglobin in infants (see Gomez, Berent, Benavides-Varela, Bion, Cattarossi, Nespor, & Mehler, 2014), I discuss these networks in Sections 4.2.2, 4.2.3, 4.3.2, and 4.3.3 as displaying *activation*.

represents a network that involves the frontal (and maybe reaching temporal) areas¹⁰, as the most dominant channels were 1, 4, 5, 7 in the left hemisphere, and 13, 16 in the right hemisphere.

Predictor weights for component 1 were analyzed in a mixed model ANOVA, with time point (300 time bins for the 30 second period) as a within subjects factor, and familiarization vowel (either /i/ or /u/) as the between subjects factor for the 16 infants. Inspection of the predictor weights for component 1 shows no significant interaction between time point and vowel of familiarization, F(299, 4186) = .34, p = .99, $\eta_p^2 = .024$, indicating that the frontal activation in component 1 was similar for infants irrespective of the vowel to which they were being familiarized. Across both familiarization vowels, the HDR reflected by these predictor weights shows a peak in activity around 10 seconds, after which activity dips around 22 seconds (see Figure 4.3b). Considering the location of this component, we suggest that this network corresponds to the learning of the contingency between the suck behavior and the sound reinforcers, which did not differ between the two familiarization vowels.

¹⁰ Relative to fMRI data, in which neural activity is measured across thousands of voxels, NIRS data has a small number of channels across which activation can be recorded. In determining the dominant loadings for each component, fMRI CPCA typically use the top 5% of component loadings (Metzak et al., 2011), which isn't possible in the current analysis (this would result in reporting only one channel). Because of this, we decided a priori to use the top 6 channels of activation (regardless of hemisphere), which results in the dominant 25% of component loadings.

Figure 4.3. Experiment 6 familiarization phase—Component 1.

a) Most strongly loaded network channels depicted in orange. **b)** HDRs for component 1, split by familiarization vowel (blue: /u/, red: /i/). Error bars denote standard error of the mean.





Component 2: Given the pattern of activation as plotted in Figure 4.4a, component 2 represents a network that involves the junction between temporal and parietal areas (and maybe reaching frontal) areas, as the most dominant channels were 6, 7, 9 in the left hemisphere, and 18, 19, 21 in the right hemisphere. As described above, we consider this to be in line with a region that reflects dorsal stream activation.

Predictor weights for component 2 were analyzed in a mixed model ANOVA, with time point (300 time bins for the 30 second period) as a within subjects factor, and familiarization vowel (either /i/ or /u/) as the between subjects factor for the 16 infants. Results of this ANOVA on the predictor weights for component 2 show a significant interaction between time point and vowel of familiarization, F(299, 4186) = 1.17, p = .029, $\eta_p^2 = .077$, indicating that the temporalparietal activation reflected in component 2 was different for infants depending on the vowel to which they were being familiarized. In /u/-familiarized infants, the HDR represented by these predictor weights showed a stronger change in activation (evidenced by the more exaggerated dips and peaks in activity) than /i/-familiarized infants. Across both familiarization vowels, the HDR reflected by these predictor weights shows a peak in activity around 12 seconds, after which activity dips around 25 seconds (see Figure 4.4b). Taking into account the general location of this component, these vowel-dependent HDRs suggest greater patterns of activation in the bilateral temporal-parietal areas during /u/-familiarization than /i/-familiarization; we suggest that this may be due to the articulatory-motor match experienced by the /u/-familiarized infants, relative to infants familiarized to /i/.

Figure 4.4. Experiment 6 familiarization phase—Component 2.

a) Most strongly loaded network channels depicted in orange. **b)** HDRs for component 2, split by familiarization vowel (blue: /u/, red: /i/). Error bars denote standard error of the mean.





4.2.2.2 Constrained Principal Component Analysis: Test phase

In the test phase, two components were extracted from the NIRS data. The percentage of variance in the NIRS data that was accounted for by the GC matrix was 3.30, and the percentages of variance in GC that were accounted for by components 1 and 2 were 5.78 and

2.49, respectively. The underlying areas of activation that correspond to each component are plotted in Figures 4.5 and 4.6; component loadings were negative for each extracted network. Once again, the average predictor weights (G loadings by subject) were plotted as a function of expected HDR time (30 seconds), and represent the response of the corresponding networks during the test phase—when the infants' memory for the two different vowel sounds was probed.

Component 1: Given the pattern of activation as plotted in Figure 4.5a, the first component extracted during the test phase represents a network that involves the junction between frontal and parietal areas (and maybe reaching temporal) areas, as the most dominant channels were 1, 4, 7 in the left hemisphere, and 15, 20, 21 in the right hemisphere.

Predictor weights for component 1were analyzed in a mixed model ANOVA, with time point (300 time bins for the 30 second period) as a within subjects factor, and familiarization vowel (either /i/ or /u/) and condition (experimental or control) as the between subjects factors for the 16 infants. Results of this ANOVA on the predictor weights for component 1 show a significant interaction between time point, vowel of familiarization, and condition, *F*(299, 3588) = 1.83, p < .001, $\eta_p^2 = .13$, indicating that the frontal-parietal activation reflected in component 1 was different for infants depending on the vowel to which they were previously familiarized and the experimental condition. Follow-up analyses were split by familiarization vowel to investigate the significant 3-way interaction. The HDR in /u/-familiarized infants who heard /u/ at test (control-/u/ infants) showed stronger patterns of activation (as evidenced by the more exaggerated dips and peaks in activity) than in infants who heard /i/ at test (experimental-/u/ infants), *F*(299,1794) = 2.12, p < .001, $\eta_p^2 = .26$ (see Figure 4.5b). For infants familiarized to /i/, there was no difference in activity in component 1 between the experimental-/i/ and control-/i/ infants, *F*(299,1794) = .71, p = .99, $\eta_p^2 = .11$ (see Figure 4.5c). Across all test conditions, the

HDR reflected by these predictor weights shows a peak in activity around 11 seconds, after which activity dips around 20 seconds. Taking into account the general location of this component, these familiarization-vowel and condition dependent activation patterns suggest greater activation of the bilateral frontal-parietal area while infants familiarized to /u/ 'remember' /u/ even after a two minute delay. This activation pattern suggests another articulatory-motor match effect, this time in a memory response.

Figure 4.5. Experiment 6 test phase—Component 1.

a) Most strongly loaded network channels depicted in orange. **b)** HDRs for component 1 /u/-familiarized infants, split by test condition (blue: /u/-experimental, light blue: /u/-control). **c)** HDRs for component 1 /i/-familiarized infants, split by test condition (red: /i/-experimental, pink: /i/-control). Error bars denote standard error of the mean.



Component 2: Given the pattern of activation as plotted in Figure 4.6a, the second component extracted during the test phase represents a network that involves the left temporal and parietal and right frontal areas, as the most dominant channels were 3, 6, 7, and 10 in the left hemisphere, and 13 and 16 in the right hemisphere.

Predictor weights for component 2 were analyzed in a mixed model ANOVA, with time point (300 time bins for the 30 second period) as a within subjects factor, and familiarization vowel (either /i/ or /u/) and condition (experimental or control) as the between subjects factors for the 16 infants. Results of this ANOVA on the predictor weights for component 2 show a significant interaction between time point, vowel of familiarization, and experimental condition, $F(299, 3588) = 1.48, p < .001, \eta_p^2 = .11$, indicating that the activation reflected in component 2 was different for infants depending on the vowel to which they were previously familiarized and their experimental condition. Follow-up analyses were split by familiarization vowel to investigate the significant 3-way interaction. In /i/-familiarized infants who heard /i/ at test (control-/i/ infants), the HDR represented by these predictor weights showed stronger activation (as evidenced by a smooth HDR, and more exaggerated dips and peaks in activity) than infants who heard /u/ at test (experimental-/i/ infants), F(299,1794) = 1.55, p < .001, $\eta_p^2 = .21$ (see Figure 4.6b). For infants familiarized to /u/, there was no difference in activity in component 2 between the experimental-/u/ and control-/u/ infants, F(299,1794)=1.06, p = .26, $\eta_p^2 = .15$ (see Figure 4.6c). Across all test conditions, the HDR reflected by these predictor weights shows a peak in activity around 12 seconds, after which activity dips around 23 seconds. These findings suggest that component 2 is detecting the /i/-memory response, much like component 1 detected the /u/-memory response.

Figure 4.6. Experiment 6 test phase—Component 2.

a) Most strongly loaded network channels depicted in orange. **b)** HDRs for component 2 /i/-familiarized infants, split by test condition (red: /i/-experimental, pink: /i/-control). **c)** HDRs for component 2 /u/-familiarized infants, split by test condition (blue: /u/-experimental, light blue: /u/-control). Error bars denote standard error of the mean.



4.2.2.3 Behavioral data: High amplitude sucks

High amplitude suck data are reported here for a point of reference for the neuroimaging data reported above. Because the subject sample size in this study is lower (n=16) than those reported in Experiment 5 (Chapter 3, n=24), statistical power was low (no ANOVAs reached significance); however, we report the same analyses as those in Experiment 5, to ensure that the mean pattern of sucking is similar to that exhibited by infants in the current experiment.

Familiarization phase: High-amplitude suck data for the 16 infants were averaged over the first 2 minutes of familiarization and the last 2 minutes of familiarization, and were split by familiarization vowel. A 2 (Phase) X 2 (Familiarization Vowel) mixed ANOVA was performed on the HAS data using the within subjects factor of Phase (average first 2 familiarization minutes vs average last 2 familiarization minutes) and the between subjects factor of Familiarization Vowel (/u/ vs /i/). There was no main effect of Phase, F(1,14) = 1.82, p = .20, $\eta_p^2 = .16$, nor was there a significant interaction between Phase and Familiarization Vowel, F(1,14) = 1.05, p =.32, $\eta_p^2 = .070$. Infants who were familiarized to /u/ decreased in the mean number of high amplitude sucks from the first 2 minutes (M = 54.87, SD = 12.66) to the last 2 minutes (M =47.06, SD = 10.38) of familiarization; infants familiarized to /i/ exhibited no change in number of high amplitude sucks from the first 2 minutes (M = 41.06, SD = 21.91) to the last 2 minutes (M = 40.00, SD = 16.62) of familiarization. The average number of sucks for the familiarization phase in infants familiarized to /u/was 51.82 (SD = 10.35), compared to infants familiarized to i/who had 41.64 (SD = 17.59) mean number of high amplitude sucks. The decrease in sucking exhibited by the /u/-familiarized infants is in line with the HAS results from Experiment 5.

Test phase: The three-minute test period was analyzed to investigate memory responses in the newborns, specifically whether HA suck behavior differed depending on the Condition (Experimental or Control) and Familiarization Vowel (/u/ or /i/). A 3 (Phase) X 2 (Condition) X 2 (Familiarization Vowel) mixed ANOVA was performed on HAS data using the within subjects factor of Phase (Test Minute 1 vs Test Minute 2 vs Test Minute 3) and the between subjects factors of Condition (Experimental vs Control) and Familiarization Vowel (/u/ vs /i/). There was no significant interaction between Phase and Condition, F(2,24) = .97, p = .40, $\eta_p^2 = .074$; nor between Phase and Familiarization vowel, F(2,24) = 2.65, p = .091, $\eta_p^2 = .18$; nor between Phase, Condition, and Familiarization Vowel, F(2,24) = .078, p = .93, $\eta_p^2 = .006$. Mean high amplitude sucks for the 3-minute test period were split by familiarization vowel (as in the CPCA results above). Experimental-/u/ infants had a higher average number of sucks (M = 53.33, SD = 14.72) compared to control-/u/ infants (M = 42.83, SD = 20.96); experimental-/i/ infants had a slightly larger average number of sucks (M = 39.58, SD = 24.06) compared to control-/i/ infants (M = 39.58, SD = 24.06) compared to control-/i/ infants (M = 39.58, SD = 24.06) compared to control-/i/ infants findings did not reach significance, it does seem that infants in the experimental conditions did exhibit a greater suck response during the test phase compared to infants in the control conditions, in line with the findings from Experiment 5.

4.2.3 Discussion

Together, these findings portray relevant areas of activation involved in the learning of and memory for vowel sounds in neonates; the networks extracted are in line with previous NIRS research showing frontal areas during learning tasks (Gervain et al., 2008), and temporal areas (Benavides-Varela et al., 2011; Gervain et al., 2008; May et al., 2011; Pena et al., 2003) in response to language presentation in neonates. Further, we provide evidence of sensorimotor influences on the neural processing of vowel sounds in each experimental phase. First, during the familiarization phase, the network represented by component 2 showed a greater change in activation during /u/-familiarization than /i/-familiarization. We suggest that the pattern of

activation in this network, which is found in the general location of a dorsal stream of processing (Dehaene-Lambertz et al., 2002, 2006; Imada et al., 2006; Perani et al., 2011), reveals the sensorimotor influences related to the matching shape of the infants' mouths while being familiarized to /u/ sounds compared to infants familiarized to /i/. In line with the suggestion that the dorsal stream is present in newborns to map articulatory information to the incoming auditory speech, we indeed find such an effect, by manipulating the infants' mouths and showing a stronger pattern of activation for a sound that matches the infants' lip shapes.

During the test phase, two networks that involved distributed activation in frontaltemporal-parietal areas (and underlying memory processes) showed different levels of activation between the 4 experimental conditions; infants who were familiarized to /u/ and tested on /u/ showed the greatest pattern of activation in network 1, while infants who were familiarized to /i/ and tested on /i/ showed the greatest pattern of activation in network 2. Together, we show that a memory response is evident in two separate networks in which activation is greatest in control rather than experimental—conditions. This is in contrast to the findings of Benavides-Varela and colleagues (2011), who showed greater patterns of activation to the novel sound stimulus during test; in their study, infants in the experimental condition exhibited greater activation in temporal-parietal and frontal areas than control infants.

Given the vowel-specific differences in component 2 during familiarization, as well as in components 1 and 2 during test, it is unlikely that the CPCA is simply extracting activation specific to *sucking behavior*; indeed, we hypothesized differences in activation to the different vowels over areas implicated in motor planning and movements. By implementing an eventrelated analysis (accounting for individual suck patterns in the model matrices), we were able to deconvolve activation during each suck burst for individual infants, and thus control for any

individual differences in sucking and sound presentation. Therefore, regardless of the amount of sucking or resulting sound presentation experienced by each infant in the different phases, CPCA was able to identify different networks, as well as some auditory-to-motor sensitive areas of activation, in Experiment 6.

Are these patterns of activation due to the *contingency* between the suck behavior and the resulting sound? If the shape of the articulators is the important factor in the differential processing of the two vowel sounds, sounds that are presented randomly while infants suck on a soother should cause similar patterns of sound processing. While we do find networks specific to learning and memory in Experiment 6 (component 1, familiarization; and components 1 and 2 during test), one might still expect to find activation over temporal-parietal areas even during non-contingent sound presentation. Indeed, the learning network represented by the first familiarization component was *not* sensitive to the vowel heard during familiarization. This suggests, in comparison to the hypothesis we advanced in Experiment 5 (Chapter 3), that infants do not show different patterns of contingency learning based on the qualities of their behavior (sucking) and the qualities of the reinforcer (vowel roundness), and thus the patterns are not due to 'contingency awareness' per se (Watson, 1966). If the second familiarization componentwhich was sensitive to the auditory-motor match—does not require contingency between suck behavior and sound presentation, a similar pattern of activation should be evident even during non-contingent sound presentation. To address this possibility, a second study was conducted in which infants' suck behavior was not reinforced with the vowel sounds; instead, the suck behavior was independent of the presentation of the vowel sounds, which were presented randomly throughout the duration of the study.

4.3. Experiment 7: Neural networks involved during speech sound presentation *not contingent* on suck behavior

The results from Experiment 6 showed a set of networks involved in the learning of and memory for simple vowel sounds, as well as a network which we suggested is involved in processing a speech sound-articulator match: the brain regions involved in the network in the familiarization phase component 2 were sensitive to the type of vowel being heard. We suggested that this network represented sensorimotor influences during the processing of speech, as the lip shape of the infants affected neural processing of sounds that matched (/u/) or mismatched (/i/) the articulator shape. However, because the presentation of the speech sounds was contingent on the sucking behavior of the infant, the possibility remains open that these findings are specific to contingency learning. In the current experiment, we investigated the networks involved during speech sound presentation while infants once again sucked on a soother; the sounds were presented randomly, not-contingent on the infants' individual sucking behavior. In this way, we were able to determine any contingent-specific effects in processing.

Because non-contingent presentation of stimuli during the high-amplitude suck procedure does not result in a memory response in a test phase, particularly in a paradigm with a fixedlength of familiarization, (Trehub & Chang, 1977; Floccia, et al. 1997), we did not expect to find evidence of a memory response in Experiment 7. As such, we only report familiarization phase data for the CPCA and HAS analyses; test phase CPCA for Experiment 7 can be found in Appendix A.

4.3.1 Method

4.3.1.1 Participants

The participants in Experiment 7 were 16 neonates (n=8 males, n=8 females) with normal hearing and no documented health problems. Infants had a mean age of 1.25 days (ranging from 0-4 days), and had an average 47.56 % English exposure while in utero as reported by parents (ranging from 0%-100%). In addition to the 16 infants included in the final analyses, data from 12 infants were not included due to falling asleep (n = 4), crying or fussiness (n = 1), experimenter error (n = 1), equipment failure (n = 2), parental interference (n = 3), and failure to obtain sufficient analyzable NIRS data (n = 1).

4.3.1.2 Stimuli & apparatus

The stimuli and apparatus in Experiment 7 were identical to those used in Experiment 6.

4.3.1.3 Procedure

The procedure used in Experiment 7 was largely the same as that in Experiment 6; the main difference was the timing of vowel sound presentation. Infants were once again randomly assigned to one of four conditions: Experimental-/u/ condition, Control-/u/ condition, Experimental-/i/ condition, or Control-/i/ condition. In the experimental-/u/ condition, infants were familiarized to /u/, and tested on /i/; in the control-/u/ condition, infants were familiarized to /u/, and tested on /i/; in the experimental-/i/ condition, infants were familiarized to /u/, and tested on /u/; in the experimental-/i/ condition, infants were familiarized to /u/, and tested on /u/; in the control-/i/ condition, infants were familiarized to /i/, and tested on /u/; in the control-/i/ condition, infants were familiarized to /i/. Instead of presenting the vowels as a reinforcer to a high amplitude suck, sounds were presented

¹¹ As mentioned above, only the familiarization phase of this experiment is reported in this chapter; test phase analyses for the neural networks exhibited by infants in these 4 conditions can be found in Appendix A.

randomly throughout the 11-minute experiment, independent of infants' suck behavior. In fact, each infant in Experiment 7 heard the same pattern of sound presentation as an infant in Experiment 6; the sound streams used in Experiment 7 were each yoked from the sound streams heard by the infants in Experiment 6 (which *were* contingent on an individual infant's suck pattern) (see Trehub & Chang, 1977).

4.3.1.4 Data analysis and preparation of matrices

Data analysis and matrix preparation was identical to that described in Experiment 6; the model matrix G was once again based on an infant's suck burst behavior—not the presentation of sounds¹².

4.3.2 Results

4.3.2.1 Constrained Principal Component Analysis: Familiarization phase

Using CPCA, two components were extracted from the NIRS data collected during the familiarization phase. The percentage of variance in the NIRS data that was accounted for by the GC matrix was 6.44, and the percentages of variance in GC that were accounted for by components 1 and 2 were 10.74 and 1.61, respectively. The underlying areas of activation that correspond to each component are plotted in Figures 4.7 and 4.8; component loadings were negative for each extracted network. As described above, the average predictor weights (G loadings by subject) were plotted as a function of expected HDR time (30 seconds), and represent the response of the corresponding network during the familiarization phase for the two different vowel sounds.

¹² A separate CPCA was conducted to investigate whether models based on sound presentation would result in meaningful activation patterns and networks; however, the resulting data and networks resembled noise, so suck behavior was used to build the model matrix in Experiment 7.

Component 1: Given the pattern of activation as plotted in Figure 4.7a, component 1 represents a network that involves the left frontal and temporal-parietal areas, as the most dominant channels were 1, 2, 3, 4, 6, and 9, all in the left hemisphere.

Predictor weights for component 1 were analyzed in a mixed model ANOVA, with time point (300 time bins for the 30 second period) as a within subjects factor, and familiarization vowel (either /i/ or /u/) as the between subjects factor for the 16 infants. Inspection of the predictor weights for component 1 showed a significant interaction between time point and vowel of familiarization, F(299, 4186) = 1.70, p < .001, $\eta_p^2 = .11$, indicating that the left frontal and temporal-parietal activation in component 1 was different for infants depending on the vowel to which they were being familiarized. In /u/-familiarized infants, the HDR represented by these predictor weights showed stronger patterns of activation (as evidenced by the more exaggerated dips and peaks in activity) than /i/-familiarized infants. In activation during /u/- familiarization, the HDR reflected a peak in activity around 12 seconds, after which activity dipped around 25 seconds; /i/-familiarization activation was much shallower and less smooth in comparison, and also peaked around 12 seconds (see Figure 4.7b). Taking into account the general location of this component, this network overlaps with the dorsal stream area from the familiarization phase component 2 in Experiment 6 above (here the network is left-lateralized); we once again suggest that this may be due to the articulatory-motor match experienced by the /u/-familiarized infants (resulting in greater activation), relative to infants familiarized to /i/. Interestingly, this activation occurs even though the sound presentation was not contingent with the suck behavior. Indeed, the fact that a separable frontal network was absent in Experiment 7 (unlike component 1 from the Experiment 6 familiarization phase) may be explained by the lack of contingency (and therefore the lack of contingency learning) between suck behavior and sound presentation.

Figure 4.7. Experiment 7 familiarization phase—Component 1.

a) Most strongly loaded network channels depicted in orange. b) HDRs for component 1, split by familiarization vowel (blue: /u/, red: /i/). Error bars denote standard error of the mean.



Component 2: Given the pattern of activation as plotted in Figure 4.8a, component 2 represents a network that involves the right frontal and some temporal and parietal areas, as the most dominant channels were 13, 14, 15, 16, 20, and 22, all in the right hemisphere.

Predictor weights for component 2 were analyzed in a mixed model ANOVA, with time point (300 time bins for the 30 second period) as a within subjects factor, and familiarization vowel (either /i/ or /u/) as the between subjects factor for the 16 infants. Results of this ANOVA

on the predictor weights for component 2 showed no significant interaction between time point and vowel of familiarization, F(299, 4186) = 0.40, p = .99, $\eta_p^2 = .027$, indicating that the right hemisphere activation reflected in component 2 was similar for infants regardless of the vowel to which they were being familiarized; activity peaked at about 15 seconds, and dipped at 26 seconds for both familiarization groups (see Figure 4.8b). Thus, unlike the left-lateralized activation in network 1 that was sensitive to the familiarization-vowel, this right-lateralized network was not familiarization-vowel-dependent.

Figure 4.8. Experiment 7 familiarization phase—Component 2.

a) Most strongly loaded network channels depicted in orange. **b)** HDRs for component 2, split by familiarization vowel (blue: /u/, red: /i/). Error bars denote standard error of the mean.





4.3.2.2 Behavioral data: High amplitude sucks

As in Experiment 6, high amplitude suck data are reported here for a point of reference for the neuroimaging data reported above. Because the subject sample size in this study is lower (n=16) than those reported in Experiment 5, Chapter 3 (n=24), statistical power was low (no ANOVAs reached significance); however, we report the same analyses as those in Experiment 5, to investigate whether the mean pattern of sucking is similar to that exhibited by infants in the current experiment. Given the fact that non-contingent presentation of sounds in HAS paradigms leads to unsuccessful learning of the contingency (see Floccia et al., 1997), we did not expect to see significant differences in suck behavior.

High-amplitude suck data for the 16 infants were averaged over the first 2 minutes of familiarization and the last 2 minutes of familiarization, and were split by familiarization vowel. A 2 (Phase) X 2 (Familiarization Vowel) mixed ANOVA was performed on the HAS data for the within subjects factor of Phase (average first 2 familiarization minutes vs average last 2 familiarization minutes) and the between subjects factor of Familiarization Vowel (/u/ vs /i/). There was no main effect of Phase, F(1,14) = .079, p=.78, $\eta_p^2 = .006$, nor was there a significant interaction between Phase and Familiarization Vowel, F(1,14) = .66, p = .43, $\eta_p^2 = .045$. Infants who were familiarized to /u/ slightly increased in the mean number of high amplitude sucks from the first 2 minutes (M = 40.94, SD = 22.64) to the last 2 minutes (M = 45.06, SD = 18.68) of familiarization; infants familiarized to /i/ slightly decreased in number of high amplitude sucks from the first 2 minutes (M = 42.81, SD = 10.00) to the last 2 minutes (M = 40.81, SD = 20.57) of familiarization. The average number of sucks for the familiarization phase in infants familiarized to /u/ was 45.06 (SD = 19.61), compared to infants familiarized to /i/ who had an average of 40.73 (SD = 14.54) high amplitude sucks.

4.3.3 Discussion

The results from Experiment 7 provide support for the findings in Experiment 6, and suggest that the sensorimotor effects seen in the familiarization phase of Experiment 6 are largely independent of the contingency of the sound presentation. In the first component extracted during the familiarization phase in Experiment 7, a left-lateralized network that included frontal areas as well as the temporal-parietal areas (which we consider to be in the dorsal stream of linguistic processing), we found vowel-specific activation patterns: /u/familiarized infants showed greater patterns of activation in this network compared to /i/familiarized infants. Regardless of the timing of sound presentation, neural activation based on an infant's suck response was sensitive to the articulatory-motor match when presented with /u/vowels compared to /i/ vowels. A second network extracted in the familiarization phase (in the right hemisphere, but over similar cortical areas to the first network) was not vowel-dependent. Thus, the left-lateralized vowel-specific activation extends those findings from Experiment 6: as has been suggested by previous research, the dorsal stream for language processing, even in newborns, may be left-lateralized (Hickok & Poeppel, 2007; Perani et al., 2011), as may be the processing of linguistic information in general (Pena et al., 2003; Dehaene-Lambertz et al., 2002; but see May et al., 2011 for evidence of bilateral activation to language in neonates).

Therefore, the results from Experiment 7 further validate the use of CPCA to analyze event-related neuroimaging data collected in NIRS paradigms. As can be seen in the first component in the familiarization phase, there is an underlying neural network that is sensitive to sensorimotor information *regardless* of whether the action in the articulators (sucking) is time-locked to the sound presentation; thus, it seems that the articulatory-motor match effects seen in Experiment 6 are more general to speech sound processing, and separate from contingency

learning effects. In fact, in a combined CPCA of the familiarization phases in Experiments 6 and 7, a largely left-lateralized component over frontal and temporal-parietal areas was also sensitive to the vowel of familiarization (with greater patterns of activation during /u/ compared to /i/ presentation), not to the level of contingency (see Appendix B for results).

4.4 General discussion

Taken together, this chapter introduces CPCA (Hunter & Takane, 2002; Metzak et al., 2011; Takane & Hunter, 2001), a relatively new method for analyzing event-related neuroimaging data, as a way to analyze NIRS data in infants, and presents data that exhibit sensorimotor-specific effects in the neural processing of speech sounds shortly after birth. By controlling for individual suck patterns (and sound presentation) using CPCA, we were able to identify areas of activation and underlying networks that were related to contingency learning (Experiment 6, Familiarization Component 1), speech sound processing (Experiment 6, Familiarization Component 2; Experiment 7, Familiarization Components 1 & 2), and memory processes (Experiment 6, Test Component 1 & 2). Further, we identified networks that were involved in and sensitive to the match (or mismatch) between perceptuo-motor system and the speech sound processing systems: areas that overlap with the dorsal stream for language processing (Dehaene-Lambertz et al., 2002, 2006; Hickok & Poeppel, 2004, 2007; Imada et al., 2006; Perani et al., 2011) exhibited greater patterns of activation during the presentation of /u/sounds (which match the shape of the infants lips as they suck on a soother) compared to activation during /i/-sound presentation. Thus, we showed differential activation in areas implicated in auditory-to-motor mapping.

The fact that we see similar auditory-to-motor activation patterns even in the absence of contingency between the sucking behavior and the sound presentation is in line with and extends

work conducted with older infants. As described earlier, Yeung & Werker (2013) used an articulatory-manipulation in 4-month-old infants during an audiovisual speech matching task, and found that oral-motor gestures interacted with performance in the task. In that design, infants were not required to suck or chew on the teether in order to receive the audiovisual stimulus, and sensorimotor effects were still evident (see also Chapter 2 in this dissertation). Therefore, although there was one network that was sensitive to the contingency between the suck behavior and the sound presentation (Experiment 6, Familiarization Component 1), the sensorimotor effects of the vowel match (versus vowel mis-match) were present over dorsal stream areas regardless of contingency.

4.4.1 Sensorimotor effect: specific to /u/?

Follow-up work is needed to further clarify the sensorimotor influences on the neural processing of the vowels we used in Experiments 6 and 7. Particularly, although we argue that the increased activation during /u/ compared to /i/ is due to the articulatory-sound-match, a study must be conducted to ensure a similar effect would be found for a lip shape that matches /i/— which would require some kind of soother that simultaneously spreads the lips. In such a study, one would hypothesize that similar cortical areas as those seen in the dorsal stream activation networks in Experiments 6 and 7 would exhibit greater patterns of activation to /i/ than to /u/. However, given the fact that HAS requires a soother that allows a range of pressure changes in the nipple (to determine the HAS threshold), a custom-designed soother would be necessary.

Could the neural activation be a result of differential processing for /u/ compared to /i/, regardless of suck behavior and the articulatory-sound-match? We think this is unlikely. First, the CPCA can largely account for these differences. All NIRS data are standardized (zero-centered) before being analyzed in CPCA, which normalizes all hemodynamic responses

between infants; any increase in activation that is shown in the predictor weights (or HDR curves) is specific to the model matrix characteristics – which in this case, were determined by suck behavior. Further, not all networks were sensitive to the articulator-sound match; only those networks that involved activation over the proposed dorsal stream (and motor areas) exhibited such effects. However, a follow-up study of the same length with the same number of overall stimuli presented while the infants do *not* suck on a soother could confirm this.

Second, for newborn infants who are less than 4 days old, the majority of their listening experience has been that encountered while in utero; the auditory system is functional by 26 weeks gestation (Eisenberg, 1976; Graven & Brown, 2008; Moore & Linthicum, 2007), and the fetus has access to a variety of auditory information: placental noise, the intrauterine heartbeat, the sound of the mother's voice, and even external sounds presented at a close distance to the mother and at a high decibel level (Busnel, Granier-Deferre, & Lecanuet, 1992; Querleu et al., 1988). In terms of the speech information (particularly the speech sound frequencies) heard in utero, fetuses as young as 27 weeks gestation show evidence of being able to discriminate simple (either monosyllabic or bisyllabic) words that differ based only on a vowel (/a/ or /i/) change (Lecanuet, Granier-Deferre, & Busnel, 1989; Shahidullah & Hepper, 1994; Weikum, Oberlander, Hensch, & Werker, 2012; Zimmer, Fifer, Kim, Rey, Chao, & Myers, 1993). Although to our knowledge, no studies have directly compared $\frac{u}{-i}$ discrimination in fetuses, the frequency range in the uterine environment includes the fundamental frequencies of the /u/ and /i/ sounds described in Section 3.2.1.2. Further, because these two vowels are common to most of the world's languages, non-native language effects compared to the in-utero-listening experience of the language spoken by the mother are improbable (as in Moon et al., 2013). Therefore, it is

unlikely that the vowel-specific effects in these experiments have arisen because of different amounts of listening experience with one type of vowel over the other.

A final point related to the sensorimotor effect found in experiments 6 and 7 requires discussion. Although we did show different patterns of neural processing in areas that overlap with the dorsal stream during /u/- (sensorimotor-auditory match) compared to /i/-presentation (sensorimotor-auditory mismatch), subsequent reports must take into account the acoustic properties of the motor behavior itself—sucking on a soother likely results in perceivable acoustic attributes that may share qualities with the rounded /u/ sounds. Newborns may be preferentially processing the /u/ sound because it shares qualities with the sound of their own suck behavior, in addition to or instead of the sensorimotor match between the shape of the lips and the rounded /u/ sound. As such, it may be these acoustic attributes that infants "match" with /u/, rather than the rounded, sensorimotor action in the suck behavior. While isolating the suck behavior (sensorimotor information) from the resulting acoustic properties (auditory information) is impossible given the experimental procedure, additional studies that measure the acoustic attributes of a newborn's suck response could help to determine whether there are similarities to the acoustic properties in /u/ compared to /i/. Such information would help support the sensorimotor claims made throughout this chapter.

4.4.2 CPCA to analyze event-related data in NIRS studies

Near-infrared Spectroscopy is a relatively new neuroimaging technique that has been used to study infant perceptual and learning capabilities (Aslin et al., in press; Gervain et al, 2011; Lloyd-Fox et al., 2010), and has already advanced our knowledge of linguistic, social, and cognitive processing in the first few months of life. However, one of the greatest limitations in implementing NIRS has been its dependence on block-designs (Aslin et al., in press; Gervain et

al, 2011; Lloyd-Fox et al., 2010). By utilizing an event-related analysis technique like Constrained Principal Component Analysis, NIRS designs will be able to integrate a larger range of experimental design paradigms.

The robust negative component loadings for each component, in each phase, in both studies require discussion. Typically, negative component loadings in a network indicate a *decrease* in activation in those channels (Metzak et al., 2011); task-negative activation is common in cognitive tasks that require effortful processing (Fox, Snyder, Vincent, Corbetta, Van Essen, & Raichle, 2005). This raises the possibility that the negative loadings found throughout Experiments 6 and 7 indicate a systematic decrease in relative oxyHb concentration, and thus, a decrease in activation in response to the task demands of the high-amplitude suck procedure (which does not negate the vowel-dependent activation effects found in Experiment 6 and 7). However, developmental researchers who use NIRS disagree about the meaning of a decrease in oxygenated hemoglobin, stating that the cause of such an 'inverted response' in infants compared to a canonical HDR is still unknown (Aslin et al., in press; Gomez, Berent, Benavides-Varela, Bion, Cattarossi, Nespor, & Mehler, 2014; Zimmerman, Roche-Labarbe, Surova, Boas, Wolf, Grant, & Franceschini, 2012). Further, the shape of the hemodynamic response differs generally between infants and adults, in that infant HDRs exhibit longer time-to-peak, as well as a significantly deeper negative undershoot (Arichi, Fagiolo, Varela, Melendez-Calderon, Allievi, Merchant, Tusor, Counsell, Burdet, Beckmann, & Edwards, 2012). As such, many infant researchers interpret this inverted response simply as evidence of 'activity', and compare the stimuli-specific differences in (de)activation patterns.

While the HDRs reported in Experiments 6 and 7 may represent deactivation due to the effortful processing required for the task, or simply an 'inverted' HDR as seen in other NIRS

studies with infants, a third possibility is that the negative component loadings may be due to the particular action of the infants during the study: because the infants are actively engaging in a task, and thus are engaging their motor systems, the deactivation across all networks may be due to the *sucking* behavior and movement. In studies on the neural activation as measured by NIRS during cognitive tasks that involve finger tapping (Kirilina, Jelzow, Heine, Niessing, Wabnitz, Bruhl, Itterman, Jacobs, & Tachtsidis, 2012) or eye movements (Wenzel, Wobst, Heekeren, Kwong, Brandt, Kohl, Obrig, Dirnagl, & Villringer, 2000), and even during gross movements of the leg (Pizza, Biallas, Wolf, Valko, & Bassetti, 2009), adult oxygenated hemoglobin patterns also show a decrease in oxygenated hemoglobin (sometimes preceded by an initial increase oxyHb, and sometimes followed by a subsequent increase in oxyHb). The undershoot of the hemodynamic response indicates hypo-oxygenation, due to the motor demands of the task. This pattern of activity is seen in each of the components extracted in Experiments 6 and 7, which were each characterized by a peak decrease in oxyHb between 10 and 15 seconds (plotted as an *increase* in Figures 4.3-4.8, but due to negative component loadings, are interpreted as a decrease in oxyHb concentration). These three possibilities must be disentangled with additional research—particularly, a study that involves similar speech sound presentation that does *not* require infants to suck would help confirm or rule out the third possibility.

A few additional points of discussion must be taken into consideration. As mentioned previously, the use of CPCA in fMRI studies takes into account activation patterns in thousands of voxels; in NIRS systems, the hemodynamic response is recorded in only a handful of channels in comparison. Therefore, the networks extracted in NIRS data, compared to those extracted using fMRI, involve relatively gross areas, and must be interpreted as such. Further, because anatomical, structural measurements of underlying neural areas were not recorded, we made the

assumption that in placing the NIRS probes across the forehead, over the ears, and reaching to the back of the skull (see Figure 4.2) that we were recording HDRs over frontal, temporal, and parietal areas; in addition, NIRS does not have the spatial resolution to identify activation in specific brain areas (such as the Spt in the temporal-parietal junction, or Broca's area in the inferior frontal gyrus). Therefore, though we assume areas of activation are largely similar across infants, structural measurements should be taken into consideration to control for this issue in future research.

4.4.3 Conclusions

The findings from Experiments 6 and 7 exhibit sensorimotor influences on the neural processing of speech information in newborn infants. By implementing Constrained Principal Component Analysis, we were able to identify separate networks involved in the high-amplitude suck procedure, and provide evidence that neonatal neural signals during speech processing are sensitive to the shape of an infant's articulators. These early existing patterns of neural activation support the previous reports of a dorsal stream of language processing, and suggest that even shortly after birth, infants map sensorimotor information from their articulators while processing language.

Chapter 5 : General discussion

5.1 Summary of experimental chapters

The study of language acquisition and speech perception across development has, in the past 30 years, continued to account for the multisensory nature of speech; adult humans use more than just auditory signals when processing speech, and it is well-established that the same is largely true for human infants (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999, 2003; Pons et al., 2009; Soto-Faraco et al., 2012). The precise characteristics of infant multisensory speech perception are still under investigation, particularly the independent versus combined roles of the auditory, visual, and sensorimotor systems. The findings presented in this dissertation provide evidence that sensorimotor information from the articulators influences the perception of auditory speech in infancy, and that these effects begin within the first days of life. In line with the abundant evidence concerning the nature of auditory and audiovisual speech perception across the first years of life, these data suggest that infants integrate broadly-specified articulatory-motor behaviors while processing speech—which result in both facilitatory and inhibitory effects.

In Chapter 2, I provided data from 6-month-old infants showing that speech perception of a non-native consonant contrast could be impeded (Experiment 3) by selectively (and temporarily) immobilizing the necessary articulators using a teething toy (Experiment 2). Importantly, this interference effect was not due to general disruption by teethers, as infants were able to perceive the contrast when a different, non-intrusive teether was used (Experiment 4). By implementing a speech contrast that is not native to English, these results suggested that experience perceiving and producing the particular speech contrast is not necessary to induce

articulatory-specific interference effects, and neither is visual access to the speech information (Kuhl & Meltzoff, 1982, 1996; Yeung & Werker, 2013).

In Chapter 3, I presented data from newborn infants using a common behavioral method—the high amplitude suck procedure—and showed that neonates are able to discriminate and exhibit memory for single vowel sounds (/u/and /i/) within the first few days of life. These results add to the existing literature on the speech perception (Bertoncini et al., 1995; Byers-Heinlein et al., 2010; DeCasper & Spence, 1986; Dehaene-Lambertz & Pena, 2001; Floccia et al., 2000; Kujala, Huotilainen, Hotakainen, Lennes, Parkkonen, Fellmen, & Naatanen, 2004; May et al., 2011; McAdams & Bertoncini, 1997; Moon et al., 2013; Pena et al., 2003; Vouloumanos & Werker, 2007) as well as learning and memory (Benavides-Varela et al., 2011; Gervain et al., 2008; Haley et al., 2006; Swain, et al, 1993) capabilities in newborn infants. However, the hypothesized result of enhanced learning of the contingency during familiarization for a vowel that matched the infants' motor movements (/u/) compared to a vowel that mismatched the movements (/i/) was only marginally supported; the reinforcing properties of the two vowel sounds did not significantly affect the infants' behavior while learning the contingency (although marginal differences occurred in the last minutes of familiarization). As seen in other studies using the HAS procedure, it is difficult to show modified behavior during a contingency learning paradigm (Floccia et al., 1997). Thus, it may be the case that vowelspecific, sensorimotor effects would not be evident in infants' suck behavior alone.

Finally, in order to investigate the possibility that sensorimotor effects on speech perception are in fact present in neonates—even if not evident in the patterns of suck behavior— Chapter 4 combined the high amplitude suck procedure with Near-Infrared Spectroscopy (NIRS), an infant-friendly neuroimaging technique (Aslin et al., in press; Gervain et al., 2011;

Lloyd-Fox et al., 2010). The data presented in Experiments 6 and 7 further illuminated the processing of vowel sounds during the HAS procedure, and identified vowel-specific (sensorimotor) effects in underlying neural networks which have been purported in infant speech processing (Dehaene-Lambertz et al., 2002, 2006; Friederici, 2011, 2012; Imada et al., 2006; Perani et al., 2011). Using Constrained Principal Component Analysis, separable networks were shown, including one that was involved in the learning of the contingency between suck behavior and sound presentation (based in frontal areas) (Experiment 6). While activation in this 'learning' network did not differ between infants in the different familiarization-vowel groups (/u/ vs /i/), a second type of network was sensitive to the two vowels (based in channels that generally covered the temporal-parietal junction): there was greater activation during /u/familiarization than during /i/-familiarization (Experiments 6 and 7) in an area that overlapped with what other researchers have called the 'dorsal stream' of language processing. I suggested that this difference in activation provided evidence of sensorimotor effects in dorsal stream activity, as proposed in dual-stream-models of language processing; the dorsal stream in adults is sensitive to articulatory-motor mapping of the auditory speech information (Campbell, 2008; Friederici, 2011; Hickok & Poeppel, 2000, 2004, 2007), and a similar stream has been identified in neonates (Perani et al., 2011) and in infants who are only a few months old (Dehaene-Lambertz et al., 2002, 2006; Imada et al., 2006).

Taken together, the results from the 7 experiments presented in this dissertation not only add to the existing evidence for sensorimotor influences on the perception of speech in humans, but also support the possibility that the infant perceptual system is sensitive to the sensorimotor compatibilities between articulatory information and auditory speech—even before they have the opportunity to learn by experience producing overt speech. Further, even without additional,
correlated information provided by visual speech information, infants can map the correspondence in their covert articulatory movements with the speech they are hearing. Therefore, although the majority of infant multisensory speech perception research to date has concerned audiovisual speech processing (Bristow et al., 2008; Kuhl & Meltzoff, 1982, 1984; Kushnerenko et al., 2008; Patterson & Werker, 1999, 2003; Pons et al., 2009)—with some studies also addressing or incorporating sensorimotor information from the articulators (Coulon et al., 2013; Kuhl & Meltzoff, 1982, 1996; Yeung & Werker, 2013)—the current research suggests that sensorimotor-auditory link exists from very early in life, without the help of the visual system.

I now revisit the questions posed in the introductory chapter, and briefly answer each using the evidence provided in earlier chapters:

1. Does the sensorimotor-auditory link require experience listening to and seeing speakers produce language? Does it exist at birth?

Considering the findings from each of the experiments, it seems that experience producing speech is not necessary for the sensorimotor-auditory link to influence speech perception in infancy, and this link does not require visual speech information to be evident in preverbal infants. Even after a few months of experience perceiving the native language (and even as infants begin to babble), sensorimotor manipulations interact with (and inhibit) the perception of sounds that infants have never before heard. Further, although research has identified sensorimotor effects on audiovisual processing of speech information in infants (Chen et al., 2004; Coulon et al., 2013; Kuhl & Meltzoff, 1982; 1996; Yeung & Werker, 2013), I show that infants are sensitive to the correspondence between the information in their articulators and the incoming auditory speech information. In line with a dorsal stream of activation during language processing (Campbell, 2008; Dehaene-Lambertz et al, 2002, 2006; Friederici, 2011; Hickok & Poeppel, 2004, 2007; Imada et al., 2006; Perani et al., 2011; Saur et al, 2008), the 'correlated' information (a mode suggested by Campbell, 2008 for audiovisual speech processing) between the shape of the articulators and the auditory speech information can affect speech processing even within the first few days of life.

2. Can covert manipulations of the articulators affect auditory speech perception in the first year of life?

The answer seems to be yes: by manipulating the shape of the mouth and the articulators in manners that are *non-speech* related (by having infants chew on teething toys or suck on pacifiers), I provided evidence of an interaction in speech perception tasks in both neural and behavioral paradigms. A similar effect of non-speech covert movements of the articulators has been shown in audiovisual speech perception in infants (Yeung & Werker, 2013). Given the plentiful evidence reported in Chapter 1 with adults (showing a speech production-perception link), and the few studies in infants using imitation designs (Chen et al., 2004; Coulon et al., 2013 Kuhl & Meltzoff, 1982; 1996) that involve more *speech*-like movements (in that the articulators aren't extrinsically manipulated), the findings from Chapters 2 through 4 validate the use of teething toys to examine early sensorimotor-auditory links in auditory speech perception (see Ito et al., 2009 for non-speech manipulations in adults).

3. If so, can this early link affect an infants' behavior? Or does it only exist in underlying neural processing, perhaps in areas that overlap with the dorsal stream of language processing?

First comparing the results of Experiments 5 to Experiments 6 and 7, the answer to this question seems to be that—at least at the very beginning of post-natal life—the sensorimotor-

auditory link only reliably affects underlying neural processing, not an infant's behavior; the activation patterns in sensorimotor integration areas were sensitive to the auditory-motor match (compared to a mis-match), while the suck behavior did not significantly differ between the two vowels in a high-amplitude sucking – contingency learning paradigm. However, it may be the case that this particular behavioral measure and contingency learning paradigm was not sensitive enough to the sensorimotor influence of the articulators to affect behavior (perhaps because of the fact that the HAS procedure combines a perceptual task with a learning task). Given the difficulty of inducing a change in infants' sucking in HAS post-baseline (Floccia et al., 1997), further studies, perhaps implementing a different experimental paradigm that only measures perception and does not involve a learning component, are necessary to examine this point.

Second, taking into account the results from Experiment 3, sensorimotor manipulations *do* affect behavior in a speech perception task in 6-month-old infants. The possibility remains that the underlying neural processing seen shortly after birth is a precursor to the behavioral effects exhibited later in development. While behavioral effects of this sensorimotor link are evident in audio*visual* speech processing tasks in newborns (Chen et al., 2004; Coulon et al., 2013), additional research is necessary to identify whether the same is true for sensorimotor-auditory speech links.

4. Is the sensorimotor-auditory speech perception link one that facilitates or improves speech processing in infants? Or can it also result in inhibitory effects?

In adults, information in the visual system can both improve speech perception (Sumby & Pollack, 1954; Navarra & Soto-Faraco, 2007) and change the perceived signal (McGurk & MacDonald, 1976); visual influences on auditory speech perception are evident in infants as well (Bristow et al., 2008; Kuhl & Meltzoff, 1982, 1984; Kushnerenko et al., 2008; Patterson &

Werker, 1999, 2003; Pons et al., 2009). Concerning sensorimotor information available from the articulators, articulatory information can also result in an improvement (or at least an increase in neural activation) during speech perception tasks in adults (Hickok et al., 2003; Hickok et al., 2009), and can also change the perceived signal (Ito et al., 2009; Sams et al., 2005; Scott et al., 2013). In infants, recent evidence suggests that articulatory information interferes with audiovisual speech matching (Yeung & Werker, 2013).

In the current report, the findings in Experiment 6 and 7 show that a matching articulatory shape with the vowel sound /u/ leads to *increased* neural activation (which may signal an 'improvement' in processing) compared to a mismatched vowel /i/, only in areas implicated in auditory-to-motor mapping. In Experiment 3 (and, importantly not in Experiments 1 or 4), the ability to perceive a speech sound distinction is *impaired* when the related articulator is selectively inhibited. Together, these results suggest that the sensorimotor effects on auditory speech perception can be bidirectional in infancy.

5.2 Implications of empirical findings and future directions

Large advances in our understanding of infant perception and learning – specifically in the realm of speech and language – began in the 1970's. As Eimas and colleagues reported (Eimas et al., 1971), infants' speech sound processing capabilities reflect categorical perception of phonemes even within the first few months of life. Since Eimas and colleagues' seminal paper, research concerning speech perception capacities in infants as they develop into native listeners has continued to grow, and, as discussed throughout this dissertation, has proceeded to account for the multidimensional, multisensory nature of infants' perceptual experiences. In this dissertation, I show that infants are prepared to link the speech they hear with the movements of their own articulators even in the first hours after birth. Due to the redundant articulatory-speech information available via the sensorimotor and auditory systems, I suggest that this early linkage may be one of the reasons infants are able to so easily process and categorize speech signals.

While the motor theories of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985), and theories of embodied cognition more generally, provide a historical context with which the work in this dissertation may align, recent advances in neuroimaging and behavioral paradigms (including those used in Experiments 1-7) have allowed more nuanced investigations of the role of the sensorimotor system in speech perception in infancy. Further, theories which posit a role for the articulatory system during the perception of speech have historically faced the challenge of integrating and accounting for the abundant evidence that preverbal infants are able to categorically perceive speech, even though these infants do not yet produce speech or babbling sounds themselves (see Ohala, 1996). By approaching the question of whether there are early links between the sensorimotor and auditory speech processing systems in a way that accounts for the early existing speech processing capabilities of infants, as well as the irrefutable role that experience plays in the development of language processing skills, the experiments in this dissertation help advance the study of language acquisition by integrating sensorimotor components into infant speech perception research.

5.2.1 Sensorimotor influences on (speech) perception—how?

How are infants linking the articulatory-motor information as they perceive speech in the tasks used in the experiments reported above? To borrow a term from the perceptual learning literature, it is possible that infants' perceptual systems are identifying the *invariance* of features across domains, or the shared qualities that exist in the articulators and the speech information (Gibson, 1969). Neonates are sensitive to amodal information, and can detect regularities in

objects and events across perceptual domains (Aldridge et al., 1999; Lewkowicz & Turkewitz, 1981; Meltzoff & Borton, 1979; Meltzoff & Moore, 1977; Slater et al., 1999; Sai, 2005). In the neonatal experiments reported in Experiments 6 and 7, infants' perceptual systems may be detecting the correspondence and perceptual invariance between the shape of their rounded lips and the rounded vowel sound, leading to greater activation to /u/ than to /i/ in the temporal-parietal cortical areas. Indeed, as mentioned earlier, Campbell (2008) suggests a similar process during audiovisual speech perception in the dorsal stream. This dorsal stream is sensitive to redundant, *correlated* (or invariant) information in the auditory and visual domains, and abstracts the redundant features from the two domains into a unified percept. In the current research, the same may be true of the auditory and articulatory invariant information: if infants perceive the invariance between the information in the auditory and articulatory systems, this would result in a difference in perception between the two vowel sounds, which I report in the speech-motor-related neural areas.

It is necessary to note here that I differentiate redundancy from amodal information in general—particularly from temporal synchrony. While temporal synchrony has been shown to be useful to infants while learning pairings of multisensory information (especially auditory and visual information—see Bahrick, Lickliter, & Flom, 2004; Gogate & Bahrick, 1998; Lewkowicz & Roder, 2012; Slater et al., 1999), here I suggest that the important factors allowing infants to link speech sounds with sensorimotor information are the redundant qualities and common attributes available in both domains. Previous research has quantified the coupling between visible speech information in the face with the acoustic qualities of speech: the motion and position of a speaker's face (including the lips, chin, and cheeks) can be estimated with great accuracy (more than 80% using principal component analysis) from the acoustic properties of the

speech signal (Vatikiotis-Bateson & Munhall, in press; Yehia, Kuratate, & Vatikiotis-Bateson, 2002). These findings highlight and quantify the redundant information available in the acoustic speech signal and the associated visible facial movements (see also Chandrasekaran et al, 2009).

In the current studies, which involve auditory and sensorimotor information (rather than auditory and visual information), the quality of 'roundness' is redundant while sucking on a soother (forcing the lips into a rounded shape) and listening to an /u/ sound (which is a rounded vowel). Although the infants were not exposed to visual speech information in any of the experiments in this dissertation, the facial movements of the infants did share redundant qualities with the heard acoustic signal. I've shown that the physical realities and similarities between the acoustic and sensorimotor information are perceivable to preverbal infants, and affect the way in which they process speech.

While temporal synchrony may be helpful in processing the relationship between sensorimotor and auditory speech information, I argue that the content of the redundancy is what is necessary for this linkage to occur. On-going work in our lab is providing evidence of content-specific audiovisual speech-matching in 6-month-olds (Danielson et al., in prep): even in the presence of temporal synchrony, infants are able to notice mismatched (and therefore, nonredundant) non-native auditory and visual speech information. These findings suggest that even though these infants have never before experienced these specific speech sounds (auditory or visual), they are able to detect a mismatch between temporally synchronous visual speech and auditory speech. Thus, in line with the argument that redundant information is necessary to link multisensory speech information, the linkage between auditory and visual speech is contentspecific; further studies must confirm that the same is true for auditory and sensorimotor speech information.

The idea of detecting invariance and redundant features across domains in speech perception may also help explain the findings from Chapter 2 – if chewing on the flat teething toy (Experiment 3) created an 'un-correlated' (or anti-correlated) feature in the shape of the tongue that mis-matched both of the sounds the infants were hearing, this may have caused the /da/ and /da/ syllables to be collapsed into a single percept. In the case where the tongue was not affected or impeded (Experiments 1 and 4), perception of the contrast was unhindered, as neither correlated *nor* un-correlated articulatory-motor information was available to the infant. Although these findings were exhibited for non-native speech sounds, suggesting that specific experience is not necessary for these links, further research is necessary to determine if these kinds of inhibitory effects do require *general* language-perception experience.

Nonetheless, I do not want to argue that infants are fully prepared to process or are fully capable of perceiving sensorimotor-auditory links at birth; experience perceiving language clearly plays an important role in developing language skills (Kuhl, 2004; Kuhl et al., 1992; Kuhl et al., 2014; Soto-Faraco et al., 2012; Werker & Tees, 1984; Yeung & Werker, 2005). However, I would like to suggest that, given a newborn infant's neural architecture which includes a dorsal stream for speech-processing, detecting the invariance across auditory and sensorimotor domains may be one of the mechanisms by which she comes to map her motor experiences onto her auditory ones. As development proceeds, an interplay between the experiences of watching and listening to people speak, practicing making speech sounds, and ultimately becoming fluent language users become linked, which may strengthen the initial sensorimotor-auditory speech links in the preverbal period, in line with a dynamic systems approach (Thelen, 1991; Thelen & Smith, 1994). At a mechanistic, neurological level, perhaps the links between the dynamic experiences in the (sensori)motor systems and the perceptual systems are strengthened via a

Hebbian mechanism (Westermann & Miranda, 2004), or by other sensorimotor learning mechanisms (Catmur, Walsh, & Heyes, 2007; Keysers & Gazzola, 2014; but see Virji-Babul, Rose, Moiseeva, & Makan, 2012, for evidence of early, experience-independent sensorimotor detection mechanism). These are speculations that must be empirically considered and tested at another time, as further research is necessary to fully document how the (non-speech-specific) articulatory-motor system, the (speech-specific) productive system, and the language perception system are integrated across development.

5.2.2 Linking multiple modalities across development: audition, vision, and motor systems

As mentioned, the majority of studies conducted to date on the multisensory nature of speech perception in infancy has concerned audiovisual speech information (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999, 2003); a handful of studies have investigated the sensorimotor interactions with audiovisual speech perception (Kuhl & Meltzoff, 1996; Coulon et al., 2013; Yeung & Werker, 2013), and others have shown that speech production patterns relate to some auditory speech perception capabilities (De Paolis et al., 2011; Majorano et al., 2014; Oller & Eilers, 1988). In contrast, this dissertation presented evidence that infant speech perception can be directly affected by sensorimotor, articulatory information. Together, these groups of studies suggest that during infant speech perception, a) the visual system interacts with the auditory system; b) the sensorimotor (articulatory) system interacts with the auditory system; and c) the three systems interact with each other. This raises the question of how the information from multiple domains is combined across development to result in the unified speech percept. Researchers have suggested that infants (and adults) are sensitive to the auditory and visual correspondences in speech because of their *intermodal* nature, based on a mapping onto articulatory information (Coulon et al., 2013; Kuhl & Meltzoff, 1982; 1984; Meltzoff & Moore,

1977; Yeung & Werker, 2013). Indeed, premotor circuits help convert sensory information from an external source during imitative learning of another's behavior (Roberts, Gobes, Murugan, Olveczky, & Mooney, 2012). This begs the question of whether information from the sensorimotor modality would be more influential than the visual modality during speech perception tasks. If so, would this interaction change across development, as infants gain experience with language perception and production? These are important questions to consider in future studies.

5.2.3 Implications for infants with orofacial anomalies and disorders

A final point of discussion concerns the implications of these findings (specifically from Chapter 2, where I presented evidence that a temporary articulatory impairment interferes with speech perception) for people with speech disorders—particularly, those with orofacial anomalies or other disorders that affect the ability to fully move the articulators. In many speech and language disorders (which affect either language comprehension and/or language production), oral-motor control is impaired (Alcock, Passingham, Watkins, & Vargha-Khadem, 2000; Dworkin, Culatta, 1985; Hill, 2001; Kumin & Adams, 2000; Stark & Blackwell, 1997). Indeed, oral-motor skills in typically-developing toddlers have been shown to be related to language production skills, where children with lower levels of oral-motor control scored lower on language production tasks (Alcock, 2006). Concerning children born with orofacial anomalies, speech production can be improved in children born with ankyloglossia (a condition in which the tongue's mobility is inhibited due to an unusually short frenulum) who receive frenuloplasty (Messner & Lalakea, 2002), and in children with macroglossia who undergo tongue-reduction surgery (Shipster, Oliver, & Morgan, 2006). These studies suggest that the general capacity for oral-motor movements may be directly linked to the development of speech production.

To date, however, few studies have concerned the relationship between oral-motor behaviors and the development of speech perception. Adults with cerebral palsy who have dysarthria, a motor speech disorder which affects articulation, exhibit lower performance on a speech discrimination task compared to age- and receptive-vocabulary-matched controls (Bishop, Brown, & Robson, 1990), and have difficulty combining auditory and visual speech into a unified percept (Siva et al., 1995). Likewise, in preschoolers who present with production errors, perception of speech is less influenced by visual speech information (Desjardins et al., 1997), suggesting that difficulty with speech production may be interfering with the perception of auditory and audiovisual speech.

Research with typically-developing humans has shown that temporarily bolstering the oral-motor system can improve performance in speech perception tasks. After undergoing a period of motor-training of the articulators (by engaging in repetitive movements of the lips and tongue), typically-developing adults exhibited 'use-induced' motor plasticity, and showed an improvement in the recognition of speech sounds in an articulator-specific way (Glenberg, Sato, & Cattaneo, 2008; Sato, Grabski, Glenberg, Brisebois, Basirat, Menard, & Cattaneo, 2011). Further, when adults with normal hearing are placed in difficult listening situations, tactile aids, in concert with speech-reading, can increase recognition of speech sounds (Sparks, Kuhl, Edmonds, & Gray, 1978).

On the other hand, many adults with speech production disorders (Naeser, Palumbo, Helm-Estabrooks, Stiassny-Eder, & Albert, 1989; Weller, 1993) and congenital conditions which cause the inability to speak (Christen, Hanefeld, Kruse, Imhauser, Ernst, & Finkenstaedt, 2000;

Lenneberg 1962; MacNeilage, Rootes, & Chase, 1967) are still able to perceive speech sounds. For example, conduction aphasia is generally characterized by the ability to maintain speech comprehension capabilities, but frequent errors in speech production occur (Benson, Sheremata, Bouchard, Segarra, Price, & Geschwind, 1973; Damasio & Damasio, 1980; Goodglass, 1992; Hickok & Poeppel, 2007), suggesting that not every production disorder results in perceptual difficulties, in line with the modulating effect of sensorimotor information on speech perception argued throughout this dissertation. Nonetheless, research into the developmental trajectory of the articulatory influences in speech disorders and impairments must be characterized; particularly, future research must consider how the physical differences in the articulatory system (orofacial anomalies) compared to underlying neural conditions (such as aphasia or dysarthria) contribute to speech perception across development; such studies will help illuminate the extent to which the sensorimotor system modulates the perception of speech.

5.3 Conclusions

The study of speech perception, particularly in infancy, has been of interest to psychologists for many years, and researchers continue to discover the ways in which perceivers of a language make sense of the speech signal. The studies presented in this dissertation have provided evidence that infants use information from their own articulators as they perceive speech, even before they speak their first words. As exciting as these initial data may be, they also demonstrate the need for developmental and cognitive psychologists, neuroscientists, and clinicians to continue to work together to more fully advance our understanding of how an infant becomes a proficient language user. At the very least, however, it seems that from the first moments of life, infants are indeed putting language in the mouth.

References

- Agnew, Z. K., McGettigan, C., Banks, B., & Scott, S. K. (2013). Articulatory movements modulate auditory responses to speech. *NeuroImage*, 73, 191-9. doi:10.1016/j.neuroimage.2012.08.020
- Alcock, K. (2006). The development of oral motor control and language. *Down Syndrome Research and Practice*, 11(1), 1-8.
- Alcock, K. J., Passingham, R. E., Watkins, K. E., & Vargha-Khadem, F. (2000). Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. *Brain and Language*, 75(1), 17-33. doi:10.1006/brln.2000.2322
- Aldridge, M. A., Braga, E. S., Walton, G. E., & Bower, T. G. R. (1999). The intermodal representation of speech in newborns. *Developmental Science*, 2(1), 42-46. doi:10.1111/1467-7687.00052
- Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology : 15*(9), 839-43. doi:10.1016/j.cub.2005.03.046
- Arichi, T., Fagiolo, G., Varela, M., Melendez-Calderon, A., Allievi, A., Merchant, N., . . .
 Edwards, A. D. (2012). Development of BOLD signal hemodynamic responses in the human brain. *NeuroImage*, *63*(2), 663-73. doi:10.1016/j.neuroimage.2012.06.054
- Aslin, R. N., Shukla, M., & Emberson, L. L. (in press). Hemodynamic correlates of cognition in human infants. *Annual Review of Psychology*.
- Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, 13(3), 99-102.

- Benavides-Varela, S., Gómez, D. M., Macagno, F., Bion, R. A. H., Peretz, I., & Mehler, J. (2011). Memory in the neonate brain. *PloS One*, 6(11).
- Benson, D. F., Sheremata, W. A., Bouchard, R., Segarra, J. M., Price, D., & Geschwind, N. (1973). Conduction aphasia: A clinicopathological study. *Archives of Neurology*, 28(5), 339-346.
- 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. doi:10.1037/0096-3445.117.1.21
- Bertoncini, J., Floccia, C., Nazzi, T., & Mehler, J. (1995). Morae and syllables: Rhythmical basis of speech representations in neonates. *Language and Speech*, *38*(4), 311-329.
- Besle, J., Fort, A., Delpuech, C., & Giard, M. -H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, 20(8), 2225-2234. doi:10.1111/j.1460-9568.2004.03670.x
- Best, C. T., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*.
- Bishop, C. W., & Miller, L. M. (2009). A multisensory cortical network for understanding speech in noise. *Journal of Cognitive Neuroscience*, 21(9), 1790-1804. doi:10.1162/jocn.2009.21118
- Bishop, D. V. M., Brown, B. B., & Robson, J. (1990). The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. *Journal of Speech and Hearing Research*, 33, 210-219.
- de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language: Journal of the Linguistic Society*

of America, 67(2), 297-319.

- 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. doi:10.1162/jocn.2009.21076
- 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. doi:10.1002/dev.20032
- Busnel, M. C., Granier-Deferre, C., & Lecanuet, J. P. (1992). Fetal audition. In G. Turkewitz (Ed.), *Developmental psychobiology* (pp. 118-134). New York Academy of Sciences: New York, NY, US.
- Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science*, *21*(3), 343-348.
- Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage*, *22*(3), 1182-94. doi:10.1016/j.neuroimage.2004.03.00
- Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, 11(12), 1110-1123.
 doi:10.1093/cercor/11.12.1110
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, 10(11), 649-658.

Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 363(1493), 1001-10. doi:10.1098/rstb.2007.2155

- Campbell, R., & Dodd, B. (1980). Hearing by eye. *Quarterly Journal of Experimental Psychology*, *32*(1), 85–99. doi:10.1080/0033555800824823
- Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Comput Biol*, 5(7), e1000436. doi:10.1371/journal.pcbi.1000436
- Chen, H. C., Vaid, J., Bortfeld, H., & Boas, D. A. (2008). Optical imaging of phonological processing in two distinct orthographies. *Experimental Brain Research*, 184(3), 427-33. doi:10.1007/s00221-007-1200-0
- Chen, X., Striano, T., & Rakoczy, H. (2004). Auditory-oral matching behavior in newborns. *Developmental Science*, 7(1), 42-47. doi:10.1111/j.1467-7687.2004.00321.x
- Christen, H. J., Hanefeld, F., Kruse, E., Imhäuser, S., Ernst, J. P., & Finkenstaedt, M. (2000). Foix--Chavany--Marie (anterior operculum) syndrome in childhood: A reappraisal of worster-drought syndrome. *Developmental Medicine & Child Neurology*, 42(02), 122-132.
- Conboy, B. T., Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2008). Event-related potential studies of early language processing at the phoneme, word, and sentence levels. *Early Language Development*, 5, 23-64.
- Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: From origin to function. *The Behavioral and Brain Sciences*, *37*(2), 177-92.
 doi:10.1017/S0140525X13000903

- Cooper, R. P., Cook, R., Dickinson, A., & Heyes, C. M. (2013). Associative (not hebbian) learning and the mirror neuron system. *Neuroscience Letters*, *540*, 28-36.
- Coulon, M., Hemimou, C., & Streri, A. (2013). Effects of seeing and hearing vowels on neonatal facial imitation. *Infancy*, *18*(5), 782-796.
- Cowan, N., Suomi, K., & Morse, P. A. (1982). Echoic storage in infant perception. *Child Development*, 53(4), 984-990.
- Damasio, H., & Damasio, A. R. (1980). The anatomical basis of conduction aphasia. *Brain : A Journal of Neurology*, *103*(2), 337-350.
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology : CB*, *19*(5), 381-385.
 doi:10.1016/j.cub.2009.01.017
- Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variancea). *The Journal of the Acoustical Society of America*, *120*(1), 407-415. doi:10.1121/1.2205133
- DeCasper, A. J., & Carstens, A. A. (1981). Contingencies of stimulation: Effects on learning and emotion in neonates. *Infant Behavior and Development*, 4(0), 19 - 35. doi:10.1016/S0163-6383(81)80004-5
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174-1176.
- DeCasper, A. J., & Sigafoos, A. D. (1983). The intrauterine heartbeat: A potent reinforcer for newborns. *Infant Behavior and Development*, 6(1), 19-25.
- DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, 9(2), 133 150.

doi: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. doi:10.1038/370292a0
- Dehaene-Lambertz, G., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport*, *12*(14), 3155-3158.
- Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, *298*(5600), 2013-2015.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., Sigman, M., &
 Dehaene, S. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences*, *103*(38), 14240-14245. doi:10.1073/pnas.0606302103
- DePaolis, R. A., Vihman, M. M., & Keren-Portnoy, T. (2011). Do production patterns influence the processing of speech in prelinguistic infants? *Infant Behavior and Development*, 34(4), 590-601. doi:10.1016/j.infbeh.2011.06.005
- Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, *45*(4), 187-203. doi:10.1002/dev.20033
- Desjardins, R. N., Rogers, J., & Werker, J. F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology*, *66*(1), 85-110.
- Dick, A. S., Solodkin, A., & Small, S. L. (2010). Neural development of networks for audiovisual speech comprehension. *Brain and Language*, *114*(2), 101 - 114. doi:10.1016/j.bandl.2009.08.005

Dworkin, J. P., & Culatta, R. A. (1985). Oral structural and neuromuscular characteristics in

children with normal and disordered articulation. *Journal of Speech and Hearing Disorders*, *50*(2), 150-156.

- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303-306.
- Eisenberg, R. B. (1976). Auditory competence in early life: The roots of communicative behavior. University Park Press Baltimore.
- Fifer, W. P., & Moon, C. (1990). Auditory experience in the fetus. In *Behavior of the fetus*. (Original work published 1990).
- Floccia, C., Christophe, A., & Bertoncini, J. (1997). High-amplitude sucking and newborns: The quest for underlying mechanisms. *Journal of Experimental Child Psychology*, 64, 175-198.
- Floccia, C., Nazzi, T., & Bertoncini, J. (2000). Unfamiliar voice discrimination for short stimuli in newborns. *Developmental Science*, *3*(3), 333-343. doi:10.1111/1467-7687.00128
- Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 17(3), 816. doi:10.1037/0096-1523.17.3.816
- Fowler, C. A., Galantucci, B., & Saltzman, E. (2003). Motor theories of perception. *The Handbook of Brain Theory and Neural Networks*, 705-707.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27), 9673-9678.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function.

Physiological Reviews, *91*(4), 1357-1392.

- Friederici, A. D. (2012). Language development and the ontogeny of the dorsal pathway. *Frontiers in Evolutionary Neuroscience*, *4*, 3. doi:10.3389/fnevo.2012.00003
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain : A Journal of Neurology*, 119(2), 593-609. doi:10.1093/brain/119.2.593
- Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology*, *61*, 191-218.
- Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature Communications*, *4*, 1490. doi:10.1038/ncomms2430
- Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, 105(37), 14222-14227. doi:10.1073/pnas.0806530105
- Gervain, J., Mehler, J., Werker, J. F., Nelson, C. A., Csibra, G., Lloyd-Fox, S., . . . Aslin, R. N. (2011). Near-infrared spectroscopy: A report from the mcdonnell infant methodology consortium. *Developmental Cognitive Neuroscience*, 1(1), 22 46. doi:10.1016/j.dcn.2010.07.004
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton-Century-Crofts New York.
- Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, 462(7272), 502-504. doi:10.1038/nature08572
- Glenberg, A. M., Sato, M., & Cattaneo, L. (2008). Use-induced motor plasticity affects the processing of abstract and concrete language. *Current Biology*, *18*(7), R290-R291.

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary

relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*(2), 133-149.

- Goodale, M. A., & Milner, A. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20-25. doi:10.1016/0166-2236(92)90344-8
- Goodglass, H. (1992). Diagnosis of conduction aphasia. In S. E. Kohn (Ed.), *Conduction Aphasia*. Lawrence Erlbaum Associates: Hillsdale, NJ.
- Gómez, D. M., Berent, I., Benavides-Varela, S., Bion, R. A., Cattarossi, L., Nespor, M., &
 Mehler, J. (2014). Language universals at birth. *Proceedings of the National Academy of Sciences of the United States of America*, 111(16), 5837-41.

doi:10.1073/pnas.1318261111

- Graven, S. N., & Browne, J. V. (2008). Auditory development in the fetus and infant. *Newborn* and Infant Nursing Reviews, 8(4), 187-193. doi:10.1053/j.nainr.2008.10.010
- Haley, D. W., Weinberg, J., & Grunau, R. E. (2006). Cortisol, contingency learning, and memory in preterm and full-term infants. *Psychoneuroendocrinology*, *31*(1), 108 - 117. doi:10.1016/j.psyneuen.2005.06.007
- Hamilton, A., Wolpert, D., & Frith, U. (2004). Your own action influences how you perceive another person's action. *Current Biology : CB*, *14*(6), 493-498.
 doi:10.1016/j.cub.2004.03.007

Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, 45(6), 393-402.

doi:http://dx.doi.org.ezproxy.library.ubc.ca/10.1016/j.jcomdis.2012.06.00

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception.

Trends in Cognitive Sciences, *4*(4), 131-138.

- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99.
 doi:10.1016/j.cognition.2003.10.011
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402. doi:10.1038/nrn2113
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-Motor interaction revealed by fMRI: Speech, music, and working memory in area spt. *Journal of Cognitive Neuroscience*, 15(5), 673-682.
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing:
 Computational basis and neural organization. *Neuron*, 69(3), 407-22.
 doi:10.1016/j.neuron.2011.01.019
- Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*, *101*(5), 2725-32. doi:10.1152/jn.91099.2008
- Hill, E. L. (2001). Non-specific nature of specific language impairment: A review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders*, 36(2), 149-171.
- Hunter, A., & Takane, Y. (2002). Constrained principal component analysis: Various applications. *Journal of Educational and Behavioral Statistics*, *27*(2), 105-145. doi:10.3102/10769986027002105
- Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor

stimuli in adult humans. NeuroImage, 29(2), 368-382.

Iacoboni, M. (2008). The role of premotor cortex in speech perception: Evidence from fMRI and rTMS. *Journal of Physiology, Paris*, *102*(1-3), 31-4.

doi:10.1016/j.jphysparis.2008.03.003

- Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: A developmental magnetoencephalography study. *NeuroReport*, 17(10), 957-962. doi:10.1097/01.wnr.0000223387.51704.89
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences*, 106(4), 1245-1248. doi:10.1073/pnas.0810063106
- Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, 37(02), 229. doi:10.1017/S0305000909990432
- Jasdzewski, G., Strangman, Wagner, J., Kwong, K. K., Poldrack, R. A., & Boas, D. A. (2003). Differences in the hemodynamic response to event-related motor and visual paradigms as measured by near-infrared spectroscopy. *NeuroImage*, 20, 479-488.
- Jusczyk, P. W. (1985). *The high-amplitude sucking technique as a methodological tool in speech perception research*. Ablex Publishing.
- Jusczyk, P. W., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology*, *23*(5), 648-654. doi:10.1037/0012-1649.23.5.648
- Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, 43(3), 253 - 291. doi:10.1016/0010-0277(92)90014-9

- Kirilina, E., Jelzow, A., Heine, A., Niessing, M., Wabnitz, H., Brühl, R., . . . Tachtsidis, I.
 (2012). The physiological origin of task-evoked systemic artefacts in functional near infrared spectroscopy. *NeuroImage*, *61*(1), 70-81. doi:10.1016/j.neuroimage.2012.02.074
- Kislyuk, D. S., Möttönen, R., & Sams, M. (2008). Visual processing affects the neural basis of auditory discrimination. *Journal of Cognitive Neuroscience*, 20(12), 2175-2184. doi:10.1162/jocn.2008.20152
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science*, *237*(4819), 1195-1197.
- Kröger, B. J., Birkholz, P., & Lowit, A. (2010). Phonemic, sensory, and motor representations in an action-based neurocomputational model of speech production. In B. Maassen & Van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research*. Oxford University Press. doi:10.1007/978-3-540-76442-7_16
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5(11), 831-43. doi:10.1038/nrn1533
- Kuhl, P. K. (2010). Brain mechanisms in early language acquisition. *Neuron*, 67(5), 713-27.doi:10.1016/j.neuron.2010.08.038
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138-1141.
- Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, 7(3), 361-381. doi:10.1016/S0163-6383(84)80050-8
- Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, *100*(4), 2425-2438. doi:10.1121/1.417951

- Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*(4209), 69-72.
- Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J. F., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.141096311
- Kuhl, P. K., Williams, K. A., & Meltzoff, A. N. (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 17(3), 829-840.
- 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, 606-608.
- 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. doi:10.1097/00001756-200409150-00018
- Kumin, L., & Adams, J. (2000). Developmental apraxia of speech and intelligibility in children with down syndrome. *Down Syndrome Quarterly*, *5*(3), 1-7.
- 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. doi:10.1073/pnas.0804275105
- Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, 7(1), 18-27. doi:10.1177/1745691611427305

Lecanuet, J. -P., & Granier-Deferre, C. (1993). Speech stimuli in the fetal environment. In Developmental neurocognition: Speech and face processing in the first year of life (pp. 237-248). Springer Netherlands. doi:10.1007/978-94-015-8234-6_20

- Lecanuet, J. P., Granier-Deferre, C., & Busnel, M. C. (1989). Differential fetal auditory reactiveness as a function of stimulus characteristics and state. *Seminars in Perinatology*, *13*(5), 421-9.
- Lenneberg, E. H. (1962). Understanding language without ability to speak: A case report. Journal of Abnormal and Social Psychology, 65(6), 419-425.
- Leroy, F., Glasel, H., Dubois, J., Hertz-Pannier, L., Thirion, B., Mangin, J. -F., & Dehaene-Lambertz, G. (2011). Early maturation of the linguistic dorsal pathway in human infants. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 31(4), 1500-1506. doi:10.1523/JNEUROSCI.4141-10.2011
- Lewkowicz, D. J. & Roder, B. (2012). Development of multisensory processing and the role of early experience. In B. Stein. (Ed.). The New Handbook of Multisensory Processes. Cambridge, MA: MIT Press, 607-626.
- Lewkowicz, D. J., & Turkewitz, G. (1981). Intersensory interaction in newborns: Modification of visual preferences following exposure to sound. *Child Development*, 827-832.
- Li, M., Kambhamettu, C., and Stone, M. (2005) Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics* **19**(6-7); 545-554.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431-461.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1 - 36. doi:10.1016/0010-0277(85)90021-6

- Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience & amp; Biobehavioral Reviews*, 34(3), 269 - 284. doi:10.1016/j.neubiorev.2009.07.008
- Locke, J. L. (1990). Structure and stimulation in the ontogeny of spoken language. Developmental Psychobiology [Dev Psychobiol] 1990 Nov.
- Locke, J. L. (2007). Bimodal signaling in infancy: Motor behavior, reference, and the evolution of spoken language. *Interaction Studies*, 8(1), 159-175.
 doi:http://dx.doi.org.ezproxy.library.ubc.ca/10.1075/is.8.1.1110
- Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, *13*(3), 110-4. doi:10.1016/j.tics.2008.11.008
- MacNeilage, P. F., Rootes, T. P., & Chase, R. A. (1967). Speech production and perception in a patient with severe impairment of somesthetic perception and motor control. *Journal of Speech and Hearing Research*, 10(3), 449-467.
- Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development*, 10(2), 179-204. doi:10.1080/15475441.2013.829740
- 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. doi:10.1016/j.cub.2009.09.06
- Massaro, D. W. (1998). Perceiving talking faces: From speech perception to a behavioral principle. Cambridge, Mass.: MIT Press
- Massaro, D. W. (2004). From multisensory integration to talking heads and language learning. InG. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes*.

MIT Press.

- Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, *106*(3), 1367-1381.
 doi:10.1016/j.cognition.2007.07.002
- 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. doi:10.3389/fpsyg.2011.00222
- McAdams, S., & Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *The Journal of the Acoustical Society of America*, *102*(5), 2945-2953. doi:doi:10.1121/1.420349
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748. doi:10.1038/264746a0
- Mehler, Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143-178.
- Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature*, 282(22), 403-404.
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*(4312), 75-78.
- Meltzoff, A. N., & Moore, M. K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, 25(6), 954. doi:10.1037/0012-1649.25.6.954
- Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of*

Neurophysiology, *56*(3), 640-662.

- Messner, A. H., & Lalakea, M. L. (2002). The effect of ankyloglossia on speech in children. Otolaryngology - Head and Neck Surgery, 127, 539-545.
- Metzak, P., Feredoes, E., Takane, Y., Wang, L., Weinstein, S., Cairo, T., . . . Woodward, T. S. (2011). Constrained principal component analysis reveals functionally connected load-dependent networks involved in multiple stages of working memory. *Human Brain Mapping*, *32*(6), 856-71. doi:10.1002/hbm.21072
- Miller, J. L., Sonies, B. C., & Macedonia, C. (2003). Emergence of oropharyngeal, laryngeal and swallowing activity in the developing fetal upper aerodigestive tract: An ultrasound evaluation. *Early Human Development*.
- Miller, L. M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the crossmodal integration of speech. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 25(25), 5884-5893. doi:10.1523/JNEUROSCI.0896-05.2005
- Minagawa-Kawai, Y., van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2011). Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cerebral Cortex*, 21(2), 254-61. doi:10.1093/cercor/bhq08
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, *6*, 414-417.
- Mochida, T., Kimura, T., Hiroya, S., Kitagawa, N., Gomi, H., & Kondo, T. (2013). Speech misperception: Speaking and seeing interfere differently with hearing. *PloS One*, 8(7), e68619. doi:10.1371/journal.pone.0068619
- Molavi, B., Yeung, H. H., Byers-Heinlein, K., & Werker, J. F. (in preparation). *HASware: A MATLAB-based experimental platform for implementing the high-amplitude (HA)*

sucking procedure.

- Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, 16, 495-500.
- 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. doi:10.1111/apa.12098
- Moore, J. K., & Linthicum Jr, F. H. (2007). The human auditory system: A timeline of development. *International Journal of Audiology*, 46(9), 460-478.
- Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 29(31), 9819-9825.
 doi:10.1523/JNEUROSCI.6018-08.2009
- Naeser, M. A., Palumbo, C. L., Helm-Estabrooks, N., Stiassny-Eder, D., & Albert, M. L. (1989).
 Severe nonfluency in aphasia. Role of the medial subcallosal fasciculus and other white matter pathways in recovery of spontaneous speech. *Brain: A Journal of Neurology*, *112*, 1-38.
- Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research*, 71(1), 4-12. doi:10.1007/s00426-005-0031-5?LI=true#page-1
- Navarra, J., Yeung, H. H., Werker, J. F., & Soto-Faraco, S. (2012). Multisensory interactions in speech perception. In B. E. Stein (Ed.), *The new handbook of multisensory processing* (pp. 435-452).

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. Journal of the Acoustical

Society of America, 99(3), 1718-1725.

- Okada, K., & Hickok, G. (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language*, *98*(1), 112 117. doi:10.1016/j.bandl.2006.04.006
- Okada, K., & Hickok, G. (2009). Two cortical mechanisms support the integration of visual and auditory speech: A hypothesis and preliminary data. *Neuroscience Letters*, 452(3), 219 -223. doi:10.1016/j.neulet.2009.01.060
- Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, 59(2), 441-449. doi:10.2307/1130323

Parker Jones, O., Seghier, M. L., Kawabata Duncan, K. J., Leff, A. P., Green, D. W., & Price, C. J. (2013). Auditory-motor interactions for the production of native and non-native speech. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 33(6), 2376-87. doi:10.1523/JNEUROSCI.3289-12.2013

- Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, 22(2), 237-247.
- Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, 6(2), 191-196.

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 of the United States of America*, 100(20), 11702-11705. doi:10.1073/pnas.1934290100

Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., . . . Friederici, A.D. (2011). Neural language networks at birth. *Proceedings of the National Academy of*

Sciences, 108(38), 16056-16061. doi:10.1073/pnas.1102991108

- Petrides, M., Cadoret, G., & Mackey, S. (2005). Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature*, *435*(7046), 1235-1238.
- Pizza, F., Biallas, M., Wolf, M., Valko, P. O., & Bassetti, C. L. (2009). Periodic leg movements during sleep and cerebral hemodynamic changes detected by NIRS. *Clinical Neurophysiology*, *120*(7), 1329-1334.
- 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. doi:10.1037/0096-1523.20.2.421
- 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*, 106(26), 10598-10602. doi:10.1073/pnas.0904134106
- Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. Annals of the New York Academy of Sciences, 1191, 62-88. doi:10.1111/j.1749-6632.2010.05444.x
- Pulvermüller, F., Huss, M., Kherif, F., Martin, F. M. D. P., Hauk, O., & Shtyrov, Y. (2006).
 Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America*, 103(20), 7865-7870.
 doi:10.1073/pnas.0509989103
- Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L., & Crèpin, G. (1988). Fetal hearing. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 28(3), 191 212. doi:10.1016/0028-2243(88)90030-5

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. Trends in Neurosciences,

21(5), 188-194.

- Roberts, T. F., Gobes, S. M., Murugan, M., Olveczky, B. P., & Mooney, R. (2012). Motor circuits are required to encode a sensory model for imitative learning. *Nature Neuroscience*. doi:10.1038/nn.3206
- Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In *Handbook of child psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/9780470147658.chpsy0202
- Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference:Evidence for intermodal perception at birth. *Infant and Child Development*, 14(1), 29-50.
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Brain Research: Cognitive Brain Research*, 23(2-3), 429-35.
 doi:10.1016/j.cogbrainres.2004.11.006
- Sato, M., Grabski, K., Glenberg, A. M., Brisebois, A., Basirat, A., Ménard, L., & Cattaneo, L. (2011). Articulatory bias in speech categorization: Evidence from use-induced motor plasticity. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior*, 47(8), 1001 1003. doi:10.1016/j.cortex.2011.03.009
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M. -S., . . . Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46), 18035-18040.
 doi:10.1073/pnas.0805234105
- Schwartz, J. L., Basirat, A., Ménard, L., & Sato, M. (2012). The perception-for-action-control theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25(5), 336-354.

- Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical Society of America*, 133(4), EL286-EL292. doi:doi:10.1121/1.4794932
- Shahidullah, S., & Hepper, P. G. (1994). Frequency discrimination by the fetus. *Early Human* Development, 36(1), 13-26. doi:10.1016/0378-3782(94)90029-9
- Shipster, C., Oliver, B., & Morgan, A. (2006). Speech and oral motor skills in children with beckwith wiedemann syndrome: Pre- and post-tongue reduction surgery. *International Journal of Speech-Language Pathology*, 8(1), 45-55. doi:10.1080/14417040500484401
- Siqueland, E. R., & DeLucia, C. A. (1969). Visual reinforcement of nonnutritive sucking in human infants. *Science*, 165(3898), 1144-1146.
- Siva, N., Stevens, E. B., Kuhl, P. K., & Meltzoff, A. N. (1995). A comparison between cerebralpalsied and normal adults in the perception of auditory-visual illusions. *The Journal of the Acoustical Society of America*, 98(5), 2983-2983. doi:10.1121/1.413907
- Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, 72(2), B11-B21.
- 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), 2387-99. doi:10.1093/cercor/bhl147
- Slater, A., Quinn, P. C., Brown, E., & Hayes, R. (1999). Intermodal perception at birth: Intersensory redundancy guides newborn infants' learning of arbitrary auditory- visual pairings. *Developmental Science*, 2(3), 333-338.
- Soto-Faraco, S., Calabresi, M., Navarra, J., Werker, J., & Lewkowicz, D. J. (2012). Development of audiovisual speech perception. *Multisensory Development*, 207-228.

- Sparks, D. W., Kuhl, P. K., Edmonds, A. E., & Gray, G. P. (1978). Investigating the MESA (multipoint electrotactile speech aid): The transmission of segmental features of speech. *The Journal of the Acoustical Society of America*, 63(1), 246-257.
- Stark, R. E. (1980). Stages of speech development in the first year of life. *Child Phonology*, *1*, 73-90.
- Stark, R. E., & Blackwell, P. B. (1997). Oral volitional movements in children with language impairments. *Child Neuropsychology*, 3(2), 81-97.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews. Neuroscience*, 9(4), 255-266. doi:10.1038/nrn2331
- Stumm, Barlow, S., Vantipalli, R., Finan, D., Estep, M., Seibel, L., . . . Carlson, J. (2005).
 Amplitude/burst dynamics of the non-nutritive suck in preterm infants. *Poster presented at the Pediatric Academic Societies Annual Meeting, Washington DC.*
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *26*(2), 212-215. doi:doi:10.1121/1.1907309
- Summerfield, Q. (1979). Use of visual information for phonetic perception. *Phonetica*, *36*(4-5), 314-331.
- Swain, I. U., Zelazo, P. R., & Clifton, R. K. (1993). Newborn infants' memory for speech sounds retained over 24 hours. *Developmental Psychology*, 29(2), 312-323.
- Takane, Y., & Hunter, M. A. (2001). Constrained principal component analysis: A comprehensive theory. *Applicable Algebra in Engineering, Communication and Computing*, 12(5), 391-419. doi:10.1007/s002000100081

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic

learning in 6-month-old infants. Cognition, 108(3), 850-5.

doi:10.1016/j.cognition.2008.05.009

- Thelen, E. (1991). Motor aspects of emergent speech: A dynamic approach. In *Biological and behavioral determinants of language development* (pp. 339-362).
- Thelen, E., & Smith, L. B. (1994). Dynamic systems theories. In *Handbook of child psychology: Volume 1: Theoretical models of human development (5th ed.).*

Tinbergen, N. (1951). The study of instinct. New York Oxford University Press.

- Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, 26(7), 952-981. doi:10.1080/01690960903498424
- Trehub, S. E., & Chang, H. -W. (1977). Speech as reinforcing stimulation for infants. Developmental Psychology, 13(2), 170. doi:10.1037/0012-1649.13.2.170
- Tsao, F. -M., Liu, H. -M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development*, 75(4), 1067-1084. doi:10.1111/j.1467-8624.2004.00726.x
- Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio–visual speech perception is special. *Cognition*, 96(1), B13 - B22. doi:10.1016/j.cognition.2004.10.004
- Vatakis, A., Ghazanfar, A. A., & Spence, C. (2008). Facilitation of multisensory integration by the "unity effect" reveals that speech is special. *Journal of Vision*, 8(9), 14.1-11. doi:10.1167/8.9.14
- Vatikiotis-Bateson, E. & Munhall, K.G. (2015, in press). Audiovisual speech processing: something doesn't add up. In M. Redford (Ed.). *Handbook of speech production*, Wiley-Blackwell, Oxford.
- Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Blackwell Publishing.
- Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, 7(3), 270-276. doi:10.1111/j.1467-7687.2004.00345.x
- Vouloumanos, A., & Werker, J. F. (2007a). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159-164.
- Vouloumanos, A., & Werker, J. F. (2007b). Why voice melody alone cannot explain neonates' preference for speech. *Developmental Science*, *10*(2), 169-171. doi:10.1111/j.1467-7687.2007.00551.x
- 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.
- Warlaumont, A. S., Westermann, G., Buder, E. H., & Oller, D. K. (2013). Prespeech motor learning in a neural network using reinforcement. *Neural Networks : The Official Journal* of the International Neural Network Society, 38, 64-75. doi:10.1016/j.neunet.2012.11.012
- Watkins, E., Strafella, P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8), 989-994. doi:10.1016/S0028-3932(02)00316-0
- Watson, J. S. (2001). Contingency perception and misperception in infancy: Some potential implications for attachment. *Bulletin of the Menninger Clinic*.
- Weikum, W. M., Oberlander, T. F., Hensch, T. K., & Werker, J. F. (2012). Prenatal exposure to antidepressants and depressed maternal mood alter trajectory of infant speech perception.
 Proceedings of the National Academy of Sciences, 109(Supplement 2), 17221-17227.

doi:10.1073/pnas.1121263109

- Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, & Werker, J.
 F. (2007). Visual language discrimination in infancy. *Science*, *316*(5828), 1159.
 doi:10.1126/science.1137686
- Weller, M. (1993). Anterior opercular cortex lesions cause dissociated lower cranial nerve palsies and anarthria but no aphasia: Foix-Chavany-Marie syndrome and automatic voluntary dissociation revisited. *Journal of Neurology*, 240(4), 199-208.
- Wenzel, R., Wobst, P., Heekeren, H. H., Kwong, K. K., Brandt, S. A., Kohl, M., . . . Villringer,
 A. (2000). Saccadic suppression induces focal hypooxygenation in the occipital cortex. *Journal of Cerebral Blood Flow and Metabolism*, 1103-1110. doi:10.1097/00004647-200007000-00010
- Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 349-355.
- Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*(5), 672-683.
- Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. *Phonological Development: Models, Research, Implications*, 285-311.
- 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. doi:10.1016/S0163-6383(84)80022-3
- Werker, J. F., & Yeung, H. H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, 9(11), 519-527.

Westermann, G., & Miranda, E. R. (2004). A new model of sensorimotor coupling in the

development of speech. *Brain and Language*, *89*(2), 393-400. doi:10.1016/S0093-934X(03)00345-6

- Williams, L., & Golenski, J. D. (1978). Infant speech sound discrimination: The effects of contingent versus noncontingent stimulus presentation. *Child Development*, 49(1), 213-217. doi:10.2307/1128611
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625-636.
- Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage*, 33(1), 316-25. doi:10.1016/j.neuroimage.2006.05.032
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, 7(7), 701. doi:doi:10.1038/nn1263
- Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003).
 Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex*, *13*(10), 1034-1043. doi:10.1093/cercor/13.10.1034
- Yehia, H. C., Kuratate, T., & Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, *30*(3), 555 568.
 doi:10.1006/jpho.2002.016
- Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, 68, 123-139.

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), 234-43. doi:10.1016/j.cognition.2009.08.010

- Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*. doi:10.1177/0956797612458802
- Zimmer, E. Z., Fifer, W. P., Kim, Y. -I., Rey, H. R., Chao, C. R., & Myers, M. M. (1993). Response of the premature fetus to stimulation by speech sounds. *Early Human Development*, 33(3), 207-215.
- Zimmermann, B. B., Roche-Labarbe, N., Surova, A., Boas, D. A., Wolf, M., Grant, P. E., & Franceschini, M. A. (2012). The confounding effect of systemic physiology on the hemodynamic response in newborns. *Advances in Experimental Medicine and Biology*, 737, 103-9. doi:10.1007/978-1-4614-1566-4_16

Appendices

Appendix A: Experiment 7 test phase—CPCA

In this supplemental analysis, we report the test phase NIRS (CPCA) data from Experiment 7. In the test phase, two components were extracted from the NIRS data; examination of the scree plot showed that two components accounted for the majority of the variance during the three-minute test period. The percentage of variance in the NIRS data that was accounted for by the GC matrix was 3.91, and the percentages of variance in GC that were accounted for by components 1 and 2 were 3.76 and 2.68, respectively. The underlying areas of activation that correspond to each component are plotted in Figures 1 and 2; component loadings were negative for each extracted network. Once again, the average predictor weights (G loadings by subject) were plotted as a function of expected HDR time (30 seconds), and represent the response of the corresponding network during the test phase—when memory for the two different vowel sounds was probed.

Component 1: Given the pattern of activation as plotted in Figure 1a, the first component extracted during the test phase represents a network that involves temporal (and some left temporal-parietal junction) areas, as the most dominant channels were 7 in the left hemisphere, and 16, 17, 19, 21, and 22 in the right hemisphere.

Predictor weights for component 1 were analyzed in a mixed model ANOVA, with time point (300 time bins for the 30 second period) as a within subjects factor, and familiarization vowel (either /i/ or /u/) and condition (experimental or control) as the between subjects factors for the 16 infants. Results of this ANOVA on the predictor weights for component 1 show a significant interaction between time point and vowel of familiarization, F(299, 3588) = 2.47, p <.001, $\eta_p^2 = .17$, indicating that the activation reflected in component 1 during the test phase was different for infants depending on the vowel to which they were previously familiarized (see Figure 1b). However, there was no significant interaction between time point, vowel of familiarization, and condition, F(299,3588) = .31, p = .99, $\eta_p^2 = .025$; nor between time point and condition, F(299,3588) = .89, p = .90, $\eta_p^2 = .069$.

Figure 1. Experiment 7 test phase—Component 1.

a) Most strongly loaded network channels depicted in orange. **b)** HDRs for component 1 split by familiarization vowel (red: /i/-familiarized, blue: /u/-familiarized). Error bars denote standard error of the mean.



Follow-up investigation of the time point by familiarization vowel interaction showed that /u/-familiarized infants exhibited greater activation during the test minutes than /i/- familiarized infants, regardless of the vowel they were currently hearing. Importantly, in line with previous research showing that memory and discrimination is not evident during non-contingent sound presentation (Floccia et al., 1997; Trehub & Chang, 1977), there were no

differences in activation between experimental and control conditions—and included no frontal activation as seen in the test phase components in Experiment 6. Taking into account the fact that the general location of this component is over temporal areas, these familiarization-vowel dependent activation patterns during the test minutes may be caused by the residual activation (or 'memory') from the familiarization vowel.

Component 2: Given the pattern of activation as plotted in Figure 2a, the second component extracted during the test phase represents a network that involves the left frontal-temporal and parietal and right parietal areas, as the most dominant channels were 3, 4, 6, and 10 in the left hemisphere, and 20 and 24 in the right hemisphere.

Predictor weights for component 2 were analyzed in a mixed model ANOVA, with time point (300 time bins for the 30 second period) as a within subjects factor, and familiarization vowel (either /i/ or /u/) and condition (experimental or control) as the between subjects factors for the 16 infants. Results of this ANOVA on the predictor weights for component 2 show a significant interaction between time point, vowel of familiarization, and experimental condition, $F(299, 3588) = 2.08, p < .001, \eta_p^2 = .14$, indicating that the activation reflected in component 2 was different for infants depending on the vowel to which they were previously familiarized and their experimental condition. Follow-up analyses were split by familiarization vowel to investigate the significant 3-way interaction. Results showed that activity in /u/-familiarized infants, regardless of experimental condition, was similar, $F(299,1794) = .90, p = .89, \eta_p^2 = .130$; the HDR curves were smooth, peaking at 15 seconds (see Figure 2b). In contrast, although /i/familiarized infants who heard /i/ at test (control-/i/ infants) showed different activation compared to infants who heard /u/ at test (experimental-/i/ infants), F(299,1794) = 1.61, p < .001, $\eta_p^2 = .21$, this activity resembles noise; there were no obvious HDR-like peaks in either /i/- familiarized group (see Figure 2c). Again, as in component 1 reported above, we suggest that these activation patterns are in line with the fact that non-contingent suck-sound presentation does not lead to any reliable memory effects in the test phase; any activation patterns seen in this component seem to be residual effects from the familiarization phase.

Figure 2. Experiment 7 test phase—Component 2.

a) Most strongly loaded network channels depicted in orange. **b)** HDRs for component 2 /u/-familiarized infants, split by test condition (blue: /u/-experimental, light blue: /u/-control). **c)** HDRs for component 2 /i/-familiarized infants, split by test condition (red: /i/-experimental, pink: /i/-control). Error bars denote standard error of the mean.



The activation seen in both components extracted during the test phase replicates the previous reports of a lack of discrimination after non-contingent sound presentation; while we were able to extract components that accounted for a similar percent of variance as seen in the

Experiments in Chapter 4, the activation seems to be resulting from previous phases of the experiment rather than exposing any test-specific effects. Thus, we suggest that the activation in the test phase components in a non-contingent paradigm largely indicates the fact that infants were being presented auditory information—as the temporal areas were the main areas of activation in both networks.

Appendix B: Combining Experiment 6 and 7 familiarization phases—CPCA

In this final supplemental analysis, the oxyHb concentration change data for both Experiment 6 (n=16) and Experiment 7 (n=16) familiarization phases were combined in one CPCA (n=32); this analysis allowed the investigation of whether the extracted networks that overlap with the 'dorsal stream' seen in Experiment 6 (component 2, familiarization) and Experiment 7 (component 1, familiarization) are in fact similar across experiments. Further, the effects of contingency of sound presentation on suck behavior could be assessed. If dorsal stream activation is not based on contingency (as argued in Chapter 4), the patterns of activation over temporal-parietal areas (and resulting HDRs) should be dependent only on the vowel of presentation. Therefore, we hypothesized that a dorsal stream network would show greater activation during the presentation of /u/ compared to /i/, regardless of the contingency.

As in Experiments 6 and 7, two components were extracted from the NIRS data; examination of the scree plot showed that two components accounted for the majority of the variance during the 6-minute familiarization periods. The percentage of variance in the NIRS data that was accounted for by the GC matrix was 6.75, and the percentages of variance in GC that were accounted for by components 1 and 2 were 11.06 and 1.83, respectively. The underlying areas of activation that correspond to each component are plotted in Figures 3 and 4; component loadings were negative for each extracted network. Once again, the average predictor weights (G loadings by subject) were plotted as a function of expected HDR time (30 seconds), and represent the response of the corresponding network during the familiarization phase of each experiment.

Component 1: Given the pattern of activation as plotted in Figure 3a, the first component extracted during the test phase represents a network that involves frontal and left

182

temporal-parietal areas, as the most dominant channels were 1, 4, 6, 7, 9 in the left hemisphere, and 13 in the right hemisphere.

Predictor weights for component 1 were analyzed in a mixed model ANOVA, with time point (300 time bins for the 30 second period) as a within subjects factor, and familiarization vowel (either /i/ or /u/) and Experiment (contingent or non-contingent) as the between subjects factors for the 32 infants. Results of this ANOVA on the predictor weights for component 1 show a significant interaction between time point and vowel of familiarization, F(299,8372) =2.32, p < .001, $\eta_p^2 = .076$, indicating that the activation reflected in component 1 during the familiarization phases was different for infants depending on the vowel to which they were previously familiarized (see Figure 3b). However, there was no significant interaction between time point, vowel of familiarization, and experiment, F(299,8372) = 1.01, p = .44, $\eta_p^2 = .035$; nor between time point and experiment, F(299,3872) = .95, p = .74, $\eta_p^2 = .033$. These results and areas of activation largely overlap with the vowel-sensitive networks extracted in Experiments 6 and 7 over the 'dorsal stream' of language processing.

Figure 3. Experiment 6 and Experiment 7 familiarization phases—Component 1.

a) Most strongly loaded network channels depicted in orange. b) HDRs for component 1 split by familiarization vowel (red: /i/-familiarized, blue: /u/-familiarized). Error bars denote standard error of the mean.



Component 2: Given the pattern of activation as plotted in Figure 4a, the second component extracted during the familiarization phases of Experiments 6 and 7 represents a network that involves the right temporal and parietal areas, and some left temporal-parietal activity, as the most dominant channels were 7 in the left hemisphere, and 16, 17, 19, 21, and 22 in the right hemisphere.

Predictor weights for component 2 were analyzed in a mixed model ANOVA, with time point (300 time bins for the 30 second period) as a within subjects factor, and familiarization

vowel (either /i/ or /u/) and Experiment (contingent or non-contingent) as the between subjects factors for the 32 infants. Results of this ANOVA on the predictor weights for component 2 show a significant interaction between time point, vowel of familiarization, and experiment, F(299,3872) = 1.54, p < .001, $\eta_p^2 = .052$, indicating that the activation reflected in component 2 was different for infants depending on the vowel to which they were previously familiarized and the contingency of speech sound presentation on their suck behavior. Follow-up analyses were split by familiarization vowel to investigate the significant 3-way interaction. Results showed that activity in /u/-familiarized infants, regardless of experiment, was similar, F(299,4186) = .86, p = .96, $\eta_p^2 = .058$; the HDR curves were smooth, peaking at 15 seconds (see Figure 4b). In contrast, /i/-familiarized infants who were presented speech sounds that were non-contingent on their suck behavior (Experiment 7 infants) showed different activation compared to /i/familiarized infants who received the speech sounds contingent on their suck behavior (Experiment 6 infants), F(299,4186) = 1.43, p < .001, $\eta_p^2 = .093$ (see Figure 4c); this difference was due to the larger initial dip in activation seen in the HDR of the /i/-non-contingent infants compared to the /i/-contingent infants.

Figure 4. Experiment 6 and Experiment 7 familiarization phases—Component 2.

a) Most strongly loaded network channels depicted in orange. **b)** HDRs for component 2 /u/-familiarized infants, split by Experiment (blue: /u/-contingent, light blue: /u/-non-contingent). **c)** HDRs for component 2 /i/-familiarized infants, split by Experiment (red: /i/-contingent, pink: /i/-non-contingent). Error bars denote standard error of the mean.



The activation seen in the components extracted when familiarization data was collapsed across both Experiment 6 and Experiment 7 largely replicated the finding of 'dorsal stream' activity found in Experiments 6 and 7; here, we found a largely left-lateralized pattern of activity in component 1, in line with previous suggestions of a left hemisphere bias in the dorsal stream (Hickok et al., 2011; Perani et al., 2011). The second component from this combined CPCA showed largely right-lateralized activity, and the difference in HDRs over the areas in this network was due to the early, larger dip in activity found in the non-contingent compared to the contingent /i/-familiarized infants. However, as in the second component in Experiment 7, there was no difference in HDR between the two familiarization vowels, replicating the results from Experiment 7.