

THE ORIGINS OF ARTICULATORY-MOTOR INFLUENCES ON
SPEECH PERCEPTION

by

HO HENNY YEUNG

B.S., Duke University, 2003
M.A., The University of British Columbia, 2006

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

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2010

© Ho Henny Yeung, 2010

ABSTRACT

Myriad factors influence perceptual processing, but “embodied” approaches assert that sensorimotor information about bodily movements plays an especially critical role. This view has precedence in speech research, where it has often been assumed that the movements of one’s articulators (i.e., the tongue, lips, jaw, etc.) are closely related to perceiving speech. Indeed, previous work has shown that speech perception is influenced by concurrent stimulation of speech motor cortex or by silently making articulatory motions (e.g., mouthing “pa”) when hearing speech sounds. Critics of embodied approaches claim instead that so-called articulatory influences are attributed to other processes (e.g., auditory imagery or feedback from phonological categories), which are also activated when making speech articulations. This dissertation explores the embodied basis of speech perception, and further investigates its ontogenetic development. Chapter 2 reports a study where adults made silent and synchronous speech-like articulations while listening to and identifying speech sounds. Results show that sensorimotor aspects of these movements (i.e., articulatory-motor information) are a robust source of perceptual modulation, independent from auditory imagery or phonological activation. Chapter 3 reports that even low-level, non-speech articulatory-motor information (i.e., holding one’s breath at a particular position in the vocal tract) can exert a subtle influence on adults’ perception of related speech sounds. Chapter 4 investigates the developmental origins of these influences, showing that low-level articulatory information can influence 4.5-month-old infants’ audiovisual speech perception. Specifically, achieving lip-shapes related to /i/ and /u/ vowels (while chewing

or sucking, respectively) is shown to disrupt infants' ability to match auditory speech information about these vowels to visual displays of talking faces. Together, these chapters show that aspects of speech processing are embodied and follow a pattern of differentiation in development. Before infants produce clear speech, links between low-level articulatory representations and speech perception are already in place. As adults, these links become more specific to sensorimotor information in dynamically coordinated articulations, but vestigial links to low-level articulatory-motor information remain from infancy.

TABLE OF CONTENTS

Abstract.....	ii
Table of Contents	iv
List of Tables.....	vi
List of Figures.....	vii
Acknowledgements	viii
Dedication	x
Co-authorship Statement.....	xi
1: Introduction	1
1.1 Theoretical orienting and basic claims.....	1
1.2 The theoretical landscape	4
1.2.1 Perceptual and cognitive processes: embodied or symbolic?	5
1.2.2 Developmental course: differentiation or enrichment?	13
1.3 Action in perception: a general overview	16
1.3.1 Historical roots.....	17
1.3.2 Motor competencies in perception of visual motion.....	18
1.3.3 Motor activation in observing and predicting actions.....	20
1.3.4 Sensorimotor information in perceptual identification.....	23
1.3.5 Developmental accounts of action-perception couplings.....	27
1.3.6 Major theoretical approaches.....	30
1.4 Action in perception: the domain of speech	33
1.4.1 Historical and theoretical roots.....	34
1.4.2 Motor activation during speech perception.....	38
1.4.3 Sensorimotor influences in perceptual analysis of speech	39
1.4.4 Developmental accounts of action-perception couplings in speech.....	41
1.5 Conclusions	45
1.6 References.....	48
2: Sensorimotor Aspects of Articulation Modulate Auditory Speech	
Perception	66
2.1 Experiment 1	70
2.1.1 Method	73
2.1.2 Results	75
2.1.3 Discussion	77
2.2 Experiment 2.....	78
2.2.1 Method	80
2.2.2 Results	81
2.2.3 Discussion	82
2.3 General Discussion	84
2.4 References.....	87

3: Maintaining a Single Articulatory Position can Influence Speech Perception	91
3.1 Method	95
3.1.1 Participants	96
3.1.2 Stimuli	97
3.1.3 Procedure	97
3.2 Results	99
3.3 Discussion	103
3.4 References	106
4: Achieving Lip-shapes While Sucking and Chewing Influences Infants' Audiovisual Speech Perception	109
4.1 Introduction	109
4.2 Experiment 1	113
4.2.1 Method	114
4.2.2 Results	117
4.2.3 Discussion	118
4.3 Experiment 2	119
4.3.1 Method	119
4.3.2 Results	120
4.3.3 Discussion	122
4.4 Experiment 3	122
4.4.1 Method	124
4.4.2 Results	126
4.4.3 Discussion	131
4.5 General Discussion	133
4.6 References	136
5: General Discussion	141
5.1 Executive summary of empirical chapters	141
5.2 Implications and directions for future research	143
5.2.1 Theoretical implications for embodied approaches	144
5.2.2 Theoretical implications for differentiation approaches	146
5.2.3 Contrast <i>versus</i> assimilation effects	151
5.3 Conclusions and broader impacts	155
5.4 References	157
Appendices	163
Appendix A	163
Experiment 1	163
Experiment 2	164
Appendix B	165

LIST OF TABLES

Table 2.1 – Predicted patterns of (mis)perception for acoustic /aba/ in Experiment 1.....	72
Table 3.1 – Predicted results listed by the corresponding breath holding condition.	96
Table 5.1 – MPis for acoustic /ava/ from Experiment 1 of Chapter 2.	164
Table 5.2 – MPis for acoustic /ava/ from Experiment 2 of Chapter 2.	164

LIST OF FIGURES

Figure 2.1 – Results for the Articulate and Imagine blocks in Experiment 1.....	77
Figure 2.2 – Results for the Articulate block in Experiment 2.....	82
Figure 3.1 – Results for perceptual identifications.	100
Figure 3.2 – Results for decision latencies.	102
Figure 4.1 – An illustration of the apparatus and procedure.....	116
Figure 4.2 – Proportion-looking at the [i]- and [u]-faces while hearing the vowel /a/ (Experiment 1).....	118
Figure 4.3 – Proportion-looking at [i]- and [u]-faces while hearing either /i/ or /u/ (Experiment 2).	121
Figure 4.4 – Still images of an infant demonstrating lip-spreading and lip- rounding.	125
Figure 4.5 – Proportion-looking at [i]- and [u]-faces while hearing either /i/ or /u/ (Experiment 3).	127
Figure 4.6 – Results showing the interaction between lip-shape and heard vowel (Experiment 3).	129
Figure 4.7 – A summary of infants' performance in Experiments 2 and 3.	131

ACKNOWLEDGEMENTS

Like any good supervisor, Janet Werker was incredibly dedicated to me, even with so many demands on her time: she read with incisive care and with superhumanly quick turnaround; she was unfailingly supportive (even to a fault) when I pursued work that had little chance of bearing fruit; and she was a fierce advocate for me with both administrators and colleagues. But beyond those qualities, Janet is set apart, in my opinion, because she is so incredibly interested in her students as real people who grow and change. Maybe this is something you realize only when you are stumbling around in her lab, as I did for several years. She was so remarkably kind to me, even when I was not the easiest of people to be around, and when even the best, most dedicated supervisors can feel only baffled and helpless. Of all the things about human development I learned from Janet, the most important lessons were not about infants at all; they were about fostering personal and professional growth in students.

This dissertation would not be possible without the sage guidance of Bryan Gick. I learned so much about this field from him, who is a boundless source of ideas, and is always willing to talk about them. Also, Mark Scott is not only an amazing collaborator, but has also been a great friend. His ideas and enthusiasm have both kept me in check and pushed me forward. His questions continue to challenge my assumptions about this work, while his anecdotes and obscure trivia keep me laughing.

Reiko Mazuka, Colin Phillips, Fei Xu, and Geoff Hall have been four polestars in my journey through the academy. Reiko was the one who took me under her wing in the first place, and I will always be indebted to her for the support, friendship, and guidance.

Colin has been a fantastic cheerleader on my behalf, and has always shown me how exciting academic work can be. Fei was a rich well of knowledge in graduate school, and was always there to make me probe more deeply, asking tough questions. Geoff was one of the best teachers I had ever had, and I am continually impressed by how the stuff I learned in his classes are the basis for just about everything I study.

I am indebted to Jeremy Biesanz, Sue Birch, Michael Chandler, Ralph Hakstian, Steven Heine, Scott Johnson, and Larry Walker for their guidance in my academic upbringing. I also want to thank colleagues that I have been blessed to have along for the ride, particularly Laurie Fais, Katie Yoshida, Judit Gervain, Mijke Rhemtulla, Krista Byers-Heinlein, Ali Greuel, Lily May, Whitney Weikum & Dilys Leung (We.H.D!), Steph Denison, Ferran(ito) Pons, Ramesh Thiruvengadaswamy, Julie Scott, Tania Zamuner, Laura Sabourin, Chandan Narayan, Kathryn Dewar, Emma Buchtel, and Travis Proulx. I also want to acknowledge all the help I've gotten from Vashti Garcia, Jessica Deglau, Jasmine Cady, Vivian Pan, Clarisa Markel, Sarah Heller, Emily Chevrier, Maria Ho, Samantha Bangayan, Julia Leibowich, and Neda Razaz-Rahmati. I have also had the pleasure of working with many talented students, who have ended up pushing me as well: Lawrence Chen, Moko Chen, Ladan Hamadani, Zarrin Hosseinzadeh, Trish Pranjivan, Raphaëlle Roy, and Jason Lau. Finally, several funding sources made this research possible: the National Science Foundation (USA), the McDonnell Foundation, and the National Science and Engineering Research Council (Canada).

My cousins, aunts, and uncles should be thanked for being ready with hot soup or a pat on the back. There are also some dear friends, without whom I could not have remained sane over these few years: thank you Jeff, Omer, Ellen, Lachlan, TJ. To my step-dad, brother, and especially my mother: I am who I am because of you. There is no better way to say I love you than just that.

DEDICATION

To memories of Dad, especially my own

CO-AUTHORSHIP STATEMENT

Chapters 2 and 3 were co-authored with Mark Scott, Bryan Gick, and Janet F. Werker. Chapter 4 was co-authored with Janet F. Werker. For Chapter 2, I took the lead in designing the research, performed the data analysis, and prepared the manuscript. I also collected the data with help from Mark Scott and Raphaëlle Roy. For Chapter 3, Mark Scott and I contributed equally to the design of the research program. I also performed the data analysis and wrote the manuscript while Mark Scott collected the data. For Chapter 4, I took the lead in designing the research, collected the data, analyzed the results, and wrote the manuscript.

1: INTRODUCTION

1.1 Theoretical orienting and basic claims

From a perspective outside that of psychology, perceptual systems in the brain are often intuited as simple recording devices. For example, as I listen to the radio, rapidly changing sound waves hit my eardrum and begin a process that involves the cochlea, the auditory nerve, etc., and which eventually leads my brain to record, and thus *perceive* these sound waves as words. At this level of description, perception describes the process of how our sense organs record information about the external world.

From a psychologist's perspective, however, perceiving the radio is not just a recording of whether and which sound waves were present. As I listen to the radio, I actively interpret and analyze sensory signals coming in from the ear. For example, I normalize the acoustic signal for the broadcaster's gender and the idiosyncratic properties of her voice. I recognize that she is speaking English, which, after many years of speaking myself, has altered the way that I can perceive certain phonetic cues. I understand that she is talking about politics and this conceptual knowledge has helped me determine that she said "vote," even if in isolation that word would have sounded more like "boat." This list of experiential and conceptual influences in perception could go on indefinitely, but the point here is that *analysis* of the sensory signal is just as important as the sensory signal itself. This is one reason that many psychologists have viewed the process of perception as so inherently interesting: a myriad of factors affect perceptual processing of the raw sensations given to us by our sense organs, and a

scientific description of these phenomena cannot be captured by physiologists or neuroscientists alone. The psychological study of perception is needed to fully describe how the phenomenological experience of perception can arise.

This dissertation makes two basic claims about perception: one is related to the basic mechanisms underlying perceptual analysis, while the other is related to the ontogeny of these mechanisms. First, the present work argues that perception is, at least in some respects, “embodied.” In other words, it is claimed that information about bodily movements and action planning are critical in understanding how sensory signals are analyzed. Second, this dissertation also describes the ontogenetic development of these embodied aspects of perception. It is claimed that this development follows a process of “differentiation,” such that links between perception and action are broadly specified in initial stages, becoming more attuned in development to more specific and functional relations between perceiving and acting. These claims are based in the domain of speech research, since both of these views have precedence in the field. On the one hand, it has often been assumed that the movements of one’s articulators (i.e., the tongue, lips, jaw, etc.) are closely related to perceiving speech (Fowler & Galantucci, 2005). On the other hand, differentiationist views have also been popular in describing the development of speech perception and production (Werker & Tees, 2005).

Several additional characteristics of the speech signal make it an excellent case study for understanding embodied aspects of perception, and for understanding differentiationist aspects of its development. First, most human perceivers have a rich base of experience perceiving and producing speech, beginning in infancy and extending into adulthood. Second, this topic has important implications for individuals impaired in either speech perception or production: children and adults with these problems often face profound difficulties in modern society. Third, there is a scientific

need for this type of research: a significant gap in literature exists with respect to research that evaluates articulatory-motor influences on speech perception. Consider, for example, Fowler & Galantucci's (2005) discussion on this point:

"In fact, evidence for motor involvement in speech perception is weak. However, apparently this is because such evidence has rarely been sought, not because many tests have yielded negative outcomes." (p. 643)

This latter observation is a particularly curious one, given the ample historical and theoretical precedent for this research. One of the first psychological theories of speech processing, the motor theory of speech perception, had assumed that articulatory processes are recruited in speech perception (Liberman, F. S. Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). Evidence quickly accumulated, which seemed to question some of the stronger versions of the motor theory, however, and this approach has not been a focus of productive speech research in the last 20 - 30 years (Lotto, Hickok, & Holt, 2009). This has changed in the last 5 years, however, as work related to the motor theory of speech perception has yielded some promising new directions. Recently published findings have suggested, for example, that transcranial magnetic stimulation (TMS) of motor cortex (D'Ausilio et al., 2009; Möttönen & Watkins, 2009; Watkins, Strafella, & Paus, 2003) or silently moving one's articulators (Ito, Tiede, & Ostry, 2009; Nasir & Ostry, 2009; Sams, Möttönen, & Sihvonen, 2005) while hearing syllables can modulate speech perception.

This dissertation further explores these recently reported phenomena in the domain of speech. Chapters 2 and 3 investigate the origin of these articulatory influences in processing, further exploring and describing the embodied basis of speech perception. Chapter 4 investigates the developmental origin of these embodied

influences on speech perception, testing pre-verbal infants. Three basic claims are made in the following empirical chapters:

- Chapter 2 describes a study where adults make silent and synchronous speech-like articulations while listening to and identifying speech sounds. Results show that sensorimotor aspects of these movements (i.e., articulatory-motor information) are a robust source of perceptual modulation, supporting an embodied approach to speech perception.
- Chapter 3 further shows that low-level articulatory-motor information embedded in non-speech gestures (i.e., simply maintaining an articulatory position when holding one's breath) can also influence adults' perception of speech sounds. However, this low-level information has a subtler effect on perception than making the dynamic speech-like movements described in Chapter 2.
- Chapter 4 shows that this low-level articulatory-motor information can also influence audiovisual speech perception from early in development, by at least 4.5 months of age. Crucially, this is an age before which infants have begun producing clear speech, and suggests that embodied links in speech may be experience-independent.

Chapter 1 is devoted to a deeper description of the research in which this dissertation is situated. Section 1.2 discusses some of the broader theoretical debates in perception that are relevant to the concepts of embodiment and differentiation. Section 1.3 discusses research from domains outside of speech (primarily vision) that investigate links between perceptual and motor processes. Section 1.4 reviews evidence related to articulatory involvement in speech perception. Finally, Section 1.5 previews the content of subsequent chapters in more detail.

1.2 The theoretical landscape

Two central, but distinct issues in cognitive science have guided a large body of research in recent years. First, researchers have conceptualized the representations or

processes involved in perception and in cognition as either *embodied* or *symbolic*.

Second, developmental research has focused on two very general classes of theories: *differentiation* versus *enrichment* accounts. This dissertation will present evidence from the domain of speech perception that will have implications for our understanding of perceptual processes vis-à-vis these two issues. What follows below are two sections that briefly outline the central differences between *embodied* versus *symbolic* approaches on the one hand, and *enrichment* versus *differentiation* approaches on the other.

1.2.1 Perceptual and cognitive processes: embodied or symbolic?

In his classic book *Vision*, which is one of the most influential texts in the study of cognitive science, David Marr defines and attempts to tackle a basic problem outlined by George Berkeley almost 300 years ago (Marr, 1982): how does the nervous system represent a complex visual scene given the limited input available from the image projected onto the retina? In the course of investigating this phenomenon, Marr made several seminal contributions to our field, including a clear exposition of how levels of analysis can affect the types of paradigms and questions used to investigate mental processes. Marr's general approach to this problem, moving from 2-D information on the retina to our rich 3-D experience, involves building a representational model. That is, the purpose of perception is to analyze sensory input in order to reconstruct a mental model (i.e., a symbolic representation) of external reality. Marr's general framework has seeded almost all of the dominant theoretical frameworks in the contemporary study of cognitive science. For example, a classical *modular* approach suggests that perceptual processing is a matter of integrating across encapsulated perceptual inputs (i.e., independent from motor processes), which output to symbolic cognitive representations, and which, in turn, construct a model of the external world (Dennett, 1986; Fodor, 1983).

This symbolic view can be contrasted with another broad view of perception and cognition: an embodied approach, which has historically been the less dominant of the two within psychology. This view assumes that perception and cognition are tied intimately to the mechanics of bodily movement, the mental simulation of action, or are radically interactive (i.e., situated) between the body and the local environment (Barsalou, 2008; Gibbs, 2006; M. Wilson, 2002). What follows below is a brief survey of the intellectual traditions that have influenced contemporary approaches to embodiment. Second, the kinds of phenomena captured by the term “embodiment” in the modern field of cognitive science are outlined. Of these, a particular sense is identified that is relevant to the work discussed in later chapters. Third, a brief discussion of *mirror* neurons is presented, which ties that literature to the study of embodiment. These sections provide some of the requisite background to understand the rather compelling effects demonstrated in the empirical chapters, or how the physical movements of one’s articulators could possibly influence the way that one perceives speech.

1.2.1.1 Intellectual traditions contributing to notions of embodiment

The philosophical roots of embodied theories, at least in Antiquity, are generally attributed to Epicurus, known more infamously for his philosophical views on social attitudes (e.g., an Epicurean feast). A common thread through both of his scientific and social philosophies, however, is a strong materialist sensibility. For example, Epicurus suggested that the “soul,” which is commonly interpreted in Greek texts to represent concepts like *consciousness* or *mental processes*, may be concentrated in certain areas of the body (i.e., inside the heart or the head), but that atomistic components of the soul are also distributed throughout, and thus inseparable from the body as a whole. This was an important precursor for materialist views in cognitive science: Epicurus was one

of the first to suggest that cognition and mental life in general might necessarily be grounded in bodily sensations.

Most rationalist philosophers in the Western tradition, including Descartes and Rousseau, have avoided this materialist bias, discussing philosophical arguments for the basic distinction between mind and body, between mental and physical life, and between subjective and objective senses of self. One of the most interesting challenges to these notions comes from Merleau-Ponty, a French philosopher writing in the mid-twentieth century. In what is generally considered his most important work, *The Phenomenology of Perception* (1962), Merleau-Ponty argues for a middle-of-the-road solution. He is not a complete materialist, but nor does he embrace a completely rationalist view of mental life. Rather, he argues, there is some meaning to be lost in always assuming detachment between the mind and body: certainly some phenomena (e.g., I see the sky) cannot be understood by detaching the *perceiver* (i.e., the bodily sense of 'I') from the objective sense of *perceiving* (i.e., the mental phenomenon of 'seeing the sky').

A similar materialist tendency is also at the root of early *psychological* theories, which eschewed formal constructions and mental abstractions in favour of motor processes and/or their mental simulations. Perhaps one of the best examples of this is William James' original ideomotor principle (1890). The phenomenon of ideomotor effects, which suggests that mental imagery of an action automatically leads to execution of that action, was not original to James. Rather, his contribution was the "psychologizing" of the relation between mental processes and bodily action. This tendency is reflected, for example, in James' discussion of emotion: recall his famous claim that we experience fear *because* we tremble, and not that we tremble because we experience fear. James was one of the first to link the kind of actions that one performs with the psychological processes involved in perception and cognition.

Finally, the perspectives generated by J.J. and Eleanor Gibson (E. J. Gibson, 1969; J. J. Gibson, 1979) make seminal contributions to contemporary views on embodiment. Several key concepts from what has been termed an “ecological” approach to (visual) perception are found in modern embodied approaches, and two of these claims are especially relevant here (a third is the basis of *differentiation* accounts, discussed below in Section 1.2.2). First, Gibsonian accounts place a priority on understanding the contents of perceptual information: what, to use their parlance, are the “objects of perception”? Ecological theories of vision claim perceptual information originates from the object directly, rather than from abstract representations reconstructed from sensory information. In this way, what we perceive is an object’s surface, rather than the fluctuations in the light that are reflected from its surface. This does away with the kind of representational infrastructure that is assumed, for example, in Marr’s approach to vision (1982).

Second, ecological views of perception emphasize that the environment *affords* only some kinds of actions for the perceiver. The goal of the perceiver is to determine *which* affordances are present in their environment, and in this way it is ultimately the potential for action that motivates perception and drives behaviour. For example, a Gibsonian might emphasize that perception of visual surfaces is driven by the possibilities in the visual environment provides for navigation, sitting, picking up objects, etc. These two points, the discarding of intermediate and abstract representations, as well as the focus on perception in guiding action, are two theoretical cornerstone on which modern embodied approaches are built.

1.2.1.2 What is meant by “embodiment” in cognitive science?

In the study of cognitive science, there is not a clean division-of-labour between phenomena that might be classified as “abstract” vs. “action-related” and symbolic vs.

embodied theories. Indeed, symbolic theories abound, even when it comes to the study of phenomena tied closely to bodily movements: for example, speech perception (Aslin, Jusczyk, & Pisoni, 1998; Massaro, 1998; Repp, 1981; Sawusch & Gagnon, 1995) or artefact-concepts and tools (Bloom, 1996). Similarly, embodied theories have also dealt with more abstract phenomena. For example, cognitive linguistic approaches seek to explain abstract concepts like *time* through our bodily understanding of space (e.g., we *move* through time, days fall *behind* us, etc.) (Lakoff & M. Johnson, 1980). Embodied approaches are also popular in explaining social phenomena, assuming that the way in which individuals empathize or take another person's perspective involves mentally simulating what it might be like to experience these emotions or situations oneself (Goldman, 2006; Meltzoff & Brooks, 2001).

Given that there are many types of behavioural phenomena described by “embodied” theories, Wilson (2002) offers a comprehensive discussion that parses the distinct meanings of this term. Two of the senses in which “embodied” might be used to describe cognitive processes are **a)** that these processes happen in the real-world (a.k.a., “situated” cognition) and **b)** that these processes happen under real-world time constraints (Chiel & Beer, 1997; Clark, 1997, 1999). In this view, almost all cognitive activities take place in the context of a continuous stream of perceptual input, motor-planning, and decision-making (e.g., talking while driving, navigating a room while conducting a visual search, hammering a nail while checking whether a portrait is level, etc.). Only a circumscribed set of cognitive activities will take place outside of the constraints of real-time, rich environmental contexts: day-dreaming or remembering, for example. Embodied approaches emphasize the futility of building representations of the external world before deciding how to act, since the environment changes so fluidly, and demands on cognitive processes occur in much faster time scales (Brooks, 1991; Clark,

1997). A third sense of the term “embodied” suggests **c)** that cognitive work is off-loaded frequently to the environment. For example, we are often able to use pen and paper to aid memory in one form of archival off-loading. Several studies have further suggested that we can use environmental cues to save ourselves cognitive work in on-line tasks as well: for example, manually rotating blocks in the game of *Tetris* to evaluate their fit is more efficient than mentally rotating these blocks before placing them (Ballard, Hayhoe, Pook, & Rao, 1997; Kirsh & Maglio, 1994). This sense of embodiment extends the scope of the cognitive system, allowing information in the environment to do “mental work.”

The two remaining senses of embodiment that Wilson (2002) refers to are **d)** that perception and memory must be understood in the way that these processes contribute to guiding action, and **e)** that cognitive processes are still tied to mechanisms of sensory processing and motor control, even when removed from real-world environments. Several studies have supported these ideas in a variety of domains. For example, perceiving others’ actions can sometimes activate motor plans in the perceiver (Rizzolatti, Fogassi, & Gallese, 2001). Other studies show that relevant actions are primed when one processes images of certain objects (i.e., objects with handles for grasping) (Chao & Martin, 2000; Handy, Grafton, Shroff, Ketay, & Gazzaniga, 2003; Tucker & Ellis, 1998) or even when one processes the meanings of action words (Glenberg, 2008; Glenberg & Kaschak, 2003; Zwaan & L. J. Taylor, 2006). Theoretical frameworks have further argued that cognitive processes, even when removed from local environments that demand some sort of action, are inherently embodied because they use mental simulations of action or sensory processes to organize information (Grush, 2004; Jeannerod, 2001).

The purpose of this section was to clarify the polysemy inherent in the term “embodied.” Various approaches have referred to embodied theories that cover many

types of cognitive phenomena, from perception to social behaviour, and all of which share some basic assumptions about the importance of acting within the context of a body that is situated in an environment. The use of the term “embodied” in this dissertation refers more precisely to the senses that are expressed when used by perception researchers (Knoblich, 2008; Prinz & Hommel, 2002; Schütz-Bosbach & Prinz, 2007): that perceptual processes are best understood through the ways in which they interface with motor-planning and action, and that they are tied to the processes of motor control in even simple laboratory tasks removed from rich environmental contexts.

1.2.1.3 Mirror neuron systems: neural evidence for embodiment?

In recent years, new evidence has suggested a neural mechanism by which actions are tied to perceptual processes. Rizzolatti and colleagues have reported specific populations of neurons in an area of rhesus monkey cortex (i.e., Brodmann’s area F5) which is homologous with a part of Broca’s area in human cortex (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti et al., 2001; Rizzolatti & Craighero, 2004). Populations of these so-called “mirror neurons” will fire *both* when a monkey observes a specific type of goal-directed action (e.g., a human or monkey hand grasping an object), and when the monkey performs the same or a similar action. This activity is selective to both the type of action (i.e., grasping or pushing), and to the temporal synchrony of the event (i.e., these neurons fire more during perception or execution of the action, as opposed to before or after the event).

These findings have motivated many theories that suggest the mirror neuron system is involved in human cognition and its evolution. Phenomena related to mirror neurons are purported to include theory of mind (Trevvarthen & Aitken, 2001), autism (Dapretto et al., 2005), and the evolution of language (Arbib, 2005; Fogassi & Ferrari, 2007; Rizzolatti & Arbib, 1998). It is difficult to show, however, that mirror neuron

systems *actually* exist in humans. Indirect evidence for this comes from studies suggesting covert facial movements (observable only with sensitive instruments measuring small amounts of electrical activity) are made when these same adults observe facial expressions of others (Dimberg, 1982; Hess & Blairy, 2001; Lhermitte, Pillon, & Serdaru, 1986; Sato & Yoshikawa, 2007). Furthermore, several studies using neurophysiological measures have suggested that sympathetic brain activity is detected when observing others' movements in both neuroimaging experiments (e.g., Rizzolatti, Fadiga, Gallese, & Fogassi, 1996) and EEG studies (Lepage & Th  ret, 2006; Oberman et al., 2005). Some theorists (Lepage & Th  ret, 2007) have also suggested that evidence for a mirror neuron system in human infants comes from studies on their rich repertoire of imitative capabilities from very early in life (Meltzoff & Moore, 1977, 1983, 1997).

While mirror systems are often invoked in descriptions of embodied phenomena, the link with speech perception is highly controversial (see Lotto et al., 2009 for discussion on this point). Currently, there is only circumstantial evidence that mirror neurons are involved in vocal communication. For example, some mirror neurons fire when a monkey performs an action (e.g., tearing sheet of paper), sees that action (e.g., a human tearing a piece of paper), *and* hears sounds produced by that action (e.g., sounds of paper-tearing) (Kohler et al., 2002). Other neurons in monkey cortex are also sensitive to mouth actions (e.g., sucking or grasping food with the lips) and have the same properties as more classical mirror neurons (Ferrari, Gallese, Rizzolatti, & Fogassi, 2003). Further work will be needed to evaluate the link between these neurons and specific vocal behaviours in monkeys, and further relate these findings to the domain of speech in humans.

1.2.2 Developmental course: differentiation or enrichment?

Enrichment accounts assume that development builds up from the initial state of infants' perceptual and conceptual systems. James (1890) offered one of the most historically influential accounts of perceptual development, which assumed an enrichment view (e.g., an infant's world was initially a "blooming, buzzing confusion."). Since then, the bulk of the perceptual development literature has argued over the character of this initial state: empiricist approaches assume that infants begin with basic learning heuristics (e.g., associations) and/or motor reactions (Cohen, Chaput, & Cashon, 2002; Elman, M. Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Munakata, McClelland, M. H. Johnson, & Siegler, 1997; Piaget, 1952), while rationalist approaches argue that the initial state involves more elaborated innate representations (Spelke, 1995; Spelke & Kinzler, 2007). What both of these approaches share, however, is the view that perceptual abilities develop incrementally, starting from some particular assumptions about the initial state, and only then building more elaborate perceptual representations of real-world objects and events.

Differentiation accounts, on the other hand, assume that both sensory input and motor interactions with the environment are rich in information content, and that mature perceptual representations are not constructed or enriched. Rather, development implies the refinement of perceptual processes towards such that they optimally processes patterns of sensory input that are both invariant (i.e., correlated across modalities) and adaptive (i.e., goal-oriented). These invariant and adaptive patterns are selected from the rich set of sensory inputs available to a learner throughout development (Thelen & Smith, 1994; E. J. Gibson & Pick, 2000). Below, two general classes of differentiation theories are briefly described. First, Gibsonian and dynamic systems theories are outlined, which emphasize the tight coupling between the environment, perception, and

action in development. Second, the phenomenon of perceptual narrowing is discussed, which similarly suggests that perceptual systems are broadly sensitive in the initial state, and are attuned or reorganized in accord with “native” perceptual categories.

1.2.2.1 Differentiation in ecological and dynamic systems theories

The Gibsonian description of development is an example *par excellence* of a differentiation account. In the initial state, rich sensory input is available as the learner explores and interacts with the environment. Perception, in other words, is rich from the beginning, and not built up from a small set of sensory primitives or innate concepts (E. J. Gibson, 1969; E. J. Gibson & Pick, 2000; J. J. Gibson & E. J. Gibson, 1955). An emphasis in these approaches is on one’s interactions with the environment in combination with the goals of the learner. One principle of learning is to be “economical,” or to detect invariant patterns across modalities in the rich perceptual input from different senses, and a second is to find patterns which “fit affordances,” or which differentiate perceptual schemas that are adaptive in a given environment (E. J. Gibson & Pick, 2000).

Dynamic systems theories also capture the essence of a Gibsonian approach to perceptual learning, but elaborate on the tight coupling between the environment, motor feedback, and the developing perceptual system (Kelso, 1995; Kugler & Turvey, 1987; Thelen & Smith, 1994). The principles of these theories can best be exemplified in the domain of motor development, where even the appearance of stage-like, Piagetian development is rarely observed. Rather, development often appears haphazard, where certain motor behaviours (e.g., the stepping reflex) can be observed at one point in development, disappear in another, and then reappear again at a later point (Thelen & D. M. Fisher, 1982). These theories explain such erratic courses of development by placing an emphasis on a dynamically changing system of constraints on behaviour that

stem from both the external environment (e.g., the surfaces on which infants will step), and from endogenous factors specific to a particular infant (e.g., the mass and weight of the legs). With experience, however, these dynamically changing constraints settle into stable, efficient patterns. These patterns may emerge slowly and in piecemeal (e.g., the ability to walk), but eventually become robust in developmental time and across many contexts (Adolph, 2000, 2008).

1.2.2.2 Perceptual reorganization and narrowing

Developmental accounts of perceptual reorganization and narrowing suggest that infants' perceptual systems are broadly sensitive in their initial states, but become more attuned to "native" patterns over the course of development (see Lewkowicz & Ghazanfar, 2009; L. S. Scott, Pascalis, & Nelson, 2007; Werker & Tees, 2005 for reviews). One of the most well described cases of this perceptual change is in the domain of speech. For example, Werker & Tees (1984) reported that English-learning infants at certain ages (i.e., 6-8 months of age and younger) show broad tuning in their perceptual sensitivity for speech: they easily discriminate a non-native phonetic contrast from Hindi. With more experience hearing English, however, older infants (i.e., 10-12 months of age) have more difficulty discriminating this non-native phonetic contrast, which is not used functionally in their native language.

Perceptual systems in young infants show some degree of organization in early stages (Eimas, Siqueland, Jusczyk, & Vigorito, 1971), but this initial state is not "universal" in the sense that the perceptual system is ready to distinguish the full set of phonetic contrasts used in the world's languages at birth (Narayan, Werker, & Beddor, 2010). Theoretical views suggest instead that there is some modest degree of organization in the initial state, but language experience serves to "reorganize" perceptual systems (Werker, 1995). The development of speech perception, for

example, shows simultaneous effects of facilitation in the perception of native phonetic categories and of decreasing sensitivity to non-native information (Werker & Tees, 2005). The set of mechanisms that shape these developmental changes in speech perception, however, is still an active area of research (Kuhl, Tsao, & Liu, 2003; Maye, Werker, & Gerken, 2002; Yeung & Werker, 2009; Yoshida, Pons, Maye, & Werker, 2010).

In summary, the purpose of this section was to briefly outline some differentiation approaches to development. What is common across these broad types of approaches (i.e., Gibsonian theories, dynamic systems theories, perceptual narrowing/reorganization approaches) is their assumption that perceptual and cognitive development in human infants begins with an initially broad sensitivity to sensory information. With development, however, sensory input that is rich in information and usually correlated across multiple modalities shapes the functioning of perceptual and cognitive systems. This often results in more developed and organized patterns of behaviour, which reflects the invariant and functional aspects of sensory experiences that are accrued from early in development.

1.3 Action in perception: a general overview

The previous section presented a brief outline of the major theoretical debates in study of perception. In this section, a more thorough review of the evidence for embodiment in perception is presented. Section 1.3.1 provides some historical background, drawing heavily from examples in vision, where the bulk of this work has been done. Section 1.3.2 reviews some empirical evidence showing that motor competencies are built into perceptual analysis of visual motion. Section 1.3.3 describes evidence that motor circuits are automatically activated when observing actions. Section 1.3.4 reports work showing how sensorimotor input can affect the analysis of perceptual

information. Section 1.3.5 describes some developmental work examining the coupling between perception and action. Finally, Section 1.3.6 describes theoretical constructs in the embodied perception literature.

1.3.1 Historical roots

The history of embodied approaches in perception is rooted in motor theories of (visual) perception, which suggested several ways that afferent feedback from muscles might contribute to visual analysis (see Viviani, 2002 for review). This idea can be traced as far back as Berkeley (1709), who suggested that depth perception arose from a ‘feeling of strain’ during accommodation (i.e., changes in the lens of the eye) and convergence (i.e., inward rotation of the eyes) when maintaining fixation on an object that moves towards or away from a perceiver. By the time the link between binocularity and depth perception was made in the late 19th century, several questions about the phenomenology of vision still remained: for example, where does the notion of size constancy come from, and how does a visual scene remain stable when one initiates saccades or programs complex eye-movements? Prevailing theories on vision at the time followed Berkeley’s precedent, and dealt with existing problems in the study of vision by assuming that the *afferent* connections, or feedback from muscles to the central nervous system, played an important role in perceptual processing (see Scheerer, 1987 for review).

In the latter half of the 19th century, a new generation of motor theories focused on *efferent*, rather than afferent connections: the role of motor *commands*, rather than the reactive feedback from the execution of actions. By the late nineteenth century, Helmholtz (1867) had asserted that simply the *intention* to execute a motor action was enough to influence perception, which came to be known as an *outflow* theory of visual perception. In Helmholtz’s view, the intentions to make saccades are incorporated with

the sensory information coming from the retina, thus compensating for movements of the whole visual field that are caused by saccades. Consider what happens when one fixates the moon behind a field of swiftly moving clouds passing from left to right. Occasionally, one might experience that the clouds themselves stay still, while the moon moves in the opposite direction as the clouds' actual movement. Helmholtz's explanation of this phenomenon was related to automatically initiated saccades that would normally follow the moving clouds (i.e., what is known as optokinetic *nystagmus*). This tendency for nystagmus is inhibited at some downstream stage, but not before the intention to make these saccades is incorporated in the calculation of how the whole visual scene would be expected to move leftwards. This information from the intention to move one's eyes is incorporated into the phenomenological experience of perception, and thus results in this illusion.

1.3.2 Motor competencies in perception of visual motion

Some of the most interesting lines of evidence in support of couplings between perceptual and motor systems have shown that physical properties of human motor movements affect perception of motion (or implied motion). Consider, for example, the "two-thirds power law," which is a lawful relation between the velocity of a moving point and the degree of curvature in its trajectory, and which is a relation that is closely obeyed when attempting to move human limbs (Lacquaniti, Terzuolo, & Viviani, 1983; Viviani & Schneider, 1991; Viviani & Terzuolo, 1982). Imagine tracing an elliptical shape on a sheet of paper: the velocity of your pencil slows as you reach the parts of the ellipse that have the most curvature. This is just one physical manifestation of the two-thirds power law.

Implicit knowledge of this motor tendency penetrates visual perception. For example, Viviani and Stucchi (1989) presented a point on a screen that moved in an

elliptical path, but did not leave a traceable line. The velocity profile of the point's movement biases the estimated eccentricity of the ellipse that is traced out (i.e., how flat or round the traced path of the point looks), and this bias closely follows the parameters established by the two-thirds power law. In cases where participants had to judge when the velocity of the point was constant while the trajectory of the path was varied, their perception was similarly biased in a way that obeyed the two-thirds power law (Viviani & Stucchi, 1992). Together, these studies show the powerful influence of implicit motor knowledge in the visual perception of motion.

A similar kind of implicit motor knowledge can also be observed in the studies examining apparent motion. Flickering two static images at various stimulus-onset asynchronies (SOAs) results in the visual perception of apparent motion, depending on the content of the two images. For example, given two pictures of an analogue clock with its hands at 12:10 or 12:20, the perceived motion will follow the shortest path between these two positions (Korte, 1915). However, when viewing an image of a person facing you with their left arm in these same clock positions, the shortest path is also a physically impossible one. A number of studies have shown that apparent motion of the human arm also follows this short path, but only at very short SOAs. When viewing these displays of apparent motion at longer (i.e., humanly possible) SOAs, the path of a human arm appears to follow a longer, physically possible path and this longer path is associated with premotor activation in the brain (Shiffrar & Freyd, 1990, 1993; Stevens, Fonlupt, Shiffrar, & Decety, 2000). Moreover, the role of the motor system is highlighted in a study showing that the body-constrained apparent motion effect was found only in limbless patients who experienced phantom sensations, but not in those who did not experience these sensations (Funk, Shiffrar, & Brugger, 2005). This data strongly

suggests that implicit knowledge about the physical movements possible by the human body can penetrate low-level visual perception.

1.3.3 Motor activation in observing and predicting actions

For many years, it has been observed that reaction times are influenced by the degree to which a response is related to the triggering stimulus (e.g., a go/no-go signal). For example, when watching dynamic stimuli that are biased towards one side (e.g., an arrow, or dot moving in a certain direction), speeded responses that are selective to the congruent side are observed (i.e., a left- or right-handed response), a phenomenon that is termed “visuomotor priming” or “stimulus-response compatibility” (Brass, Bekkering, & Prinz, 2001; Proctor & Vu, 2006; Wohlschläger & Bekkering, 2002). Speeded responses are also observed when viewing objects that afford some relevant actions (i.e., grasping responses can be facilitated when viewing graspable objects) (Craighero, Fadiga, Rizzolatti, & Umiltà, 1999; Edwards, Humphreys, & Castiello, 2003; Tucker & Ellis, 1998; Vogt, P. Taylor, & Hopkins, 2003). When viewing these dynamic events or graspable objects, motor-related activation is enhanced in the brain (Jeannerod, Arbib, Rizzolatti, & Sakata, 1995). For example, viewing graspable objects has a special influence on visual attention, often times activating attention- and motor-related areas in cortex (Chao & Martin, 2000; Grèzes & Decety, 2002; Handy et al., 2005; Handy et al., 2003).

This motor activation when observing actions may be selectively guided by the amount of motor experience that one has producing these actions. For example, Hauseisen & Knöesche (2001) showed significant activation in primary motor cortex using magnetoencephalographic (MEG) techniques as experienced pianists heard recordings of piano recitals. Comparatively less activation was found in choir singers, who have significant amounts of experience hearing piano music, but not playing it. Similar motor-related brain networks are activated in piano players (but not

inexperienced controls) while simply watching videos of finger movements performed by another pianist (Haslinger et al., 2004, 2005). Finally, several neuroimaging studies of highly trained ballet dancers and capoeiristas (i.e., practitioners of a Brazilian martial art) show greater degrees of motor-related cortical activation when these dancers and martial artists view movements that they themselves have expertise in performing (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005). Moreover, familiarity is not a confounding factor in these studies on dancers' expertise: male ballet dancers have more experience observing female-specific ballet moves, but show greater activation in motor areas when observing male-specific ballet moves than female-specific ones (Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006). This research strongly suggests that motor areas are actively modulated and recruited when observing actions that one has experience performing.

An interesting exception to this positive correlation between motor experience and increased motor activation in perception comes from a study that presented experimental participants with pictures of objects that one commonly has lots of experience grasping (i.e., doorknobs), compared to pictures of objects that participants had relatively less experience grasping (i.e. climbing holds used on artificial rock walls) (Handy, Tipper, Borg, Grafton, & Gazzaniga, 2006). In these cases, more experience grasping the object (i.e., the doorknobs, or in the case of experienced climbers, doorknobs *and* climbing holds) lead to a *decrease* in activation within cortical motor circuits when observing these objects. However, this may be a case of an exception that proves the rule: one possibility is that participants engaged in relatively less internal simulation when viewing familiar objects like doorknobs, compared to when viewing unfamiliar objects, like climbing holds. This may be because they recognized the potential for action when seeing climbing holds, but engaged in more (or more effortful)

internal simulation precisely *because* these objects were more unfamiliar. Similarly, when observing dynamic actions, as was done in neuroimaging studies on expert pianists and dancers in the studies mentioned previously, motor simulation might be obligatory, given the dynamically varying and compelling nature of the stimuli. Future work will need to further parse the complex interactions between experience, and the kinds of motor activation that is associated with perception.

When we engage motor processes upon viewing objects that afford action, this may allow us to execute or perform actions (i.e., grasping that object) more quickly and effortlessly. But what other functions might internal motor simulations, or shared information between perception and action serve? One possibility is that motor information facilitates prediction when anticipating the outcomes of an observed action. Evidence for this position comes from studies that show predictions are enhanced when participants observe their own behaviour (which is presumably easier to simulate, or which activates common representations in action and perception to a greater degree), compared to the behaviour of another person. Several studies have shown, for example, that predicting the outcomes or future courses of actions, as when dart-throwing (Knoblich & Flach, 2001) or observing handwriting movements (Knoblich, Seigerschmidt, Flach, & Prinz, 2002), are enhanced when participants view their own movements, compared to those generated by another person.

Consider an example of this predictive ability that uses Fitt's Law, which describes a trade-off between speed and accuracy in human motor performance using a variety of effectors (i.e., the finger, arm, and head) (Plamondon & Alimi, 1997). Fitt's Law established that the minimum time that it takes humans to move between two targets is determined solely by the width between targets and the "amplitude" of the movement (i.e., the trade-off between force and accuracy when programming a motor movement).

More interestingly, the time to complete a task as predicted by Fitt's Law applies not only to *imagined* movements of oneself (Decety & Jeannerod, 1995), but also to observation of someone else's movements (i.e., guessing how long it takes for a person [or robot] to move between two targets at a particular speed) (Grosjean, Shiffrar, & Knoblich, 2007). Subsequent reports have tested a neurologically impaired individual who could not move his arm fast enough to complete a motor task that unimpaired individuals normally perform in accordance with Fitt's Law. Interestingly, this clinical patient's judgements of whether another person could accomplish a similar task (when watching a video) deviated from control participants' normally observed adherence to Fitt's Law (Knoblich et al., 2002). These results strongly suggest that mental simulation of movement is involved when making predictions about the outcome or possibility of completing an action.

1.3.4 Sensorimotor information in perceptual identification

Several lines of evidence have further suggested that the sensorimotor system may *also* play a key role in the perceptual analysis of visual scenes and auditory signals. This may occur at both higher-levels of analysis (i.e., judgments about the nature of a task or the identity of individuals who create these perceptual events) and lower-levels (i.e., perceptual identification of ambiguous or multistable percepts, as well as of stimuli embedded in noise).

For example, engaging the motor system can influence judgments about the *nature* of an event. Some studies asked participants to view videos of another person lifting a box and judge how much weight was lifted based on the dynamic characteristics of the lifting motion (Bingham, 1987; Runeson & Frykholm, 1981). Interestingly, when participants were asked to lift a weight themselves while making these judgments, their estimates were consistently skewed (Hamilton, Wolpert, & Frith, 2004). Lifting a weight

oneself lead to a *contrastive* bias in estimating the amount of weight lifted in the video: for example, lifting a lighter weight biases judgments in such a way that perceivers *overestimate* the weight lifted by the agent in the video. Further researching using this weight-judgment paradigm has yielded even more interesting patterns of data. Clinical patients who have lost all sense of cutaneous touch and peripheral proprioception were also tested in this weight-judgment paradigm (Bosbach, Cole, Prinz, & Knoblich, 2005). Results showed, on the one hand, that there were no differences between control and clinical populations in estimating the amount of weight lifted in the video. However, another set of videos was filmed where the person in the video was either told the correct or incorrect weight to be lifted. When these patients were then asked to judge whether the agent in the video *knew the correct weight*, performance was at chance, and far below the performance of controls. These results suggest that there are two distinct processes at work: one process uses visual kinematic information about the lifting movements to predict the weight lifted in the video, while another process that yields information about the actor's expectations about the amount of weight to be lifted requires a motor simulation.

A similar kind of advantage in identifying or simulating actions that one can generate oneself can be found in studies of biological motion, which use video displays of point-lights attached to the major joints of the body (isolating cues to kinematic motion from other perceptual cues) (Johansson, 1973). Recent evidence has suggested that differences in perceiving friends' movements (with whom we have more visual experience) compared to displays of oneself (with whom we have more motor experience) is detectable in some tasks (Loula, Prasad, Harber, & Shiffrar, 2005). Similarly, this boost in self-recognition is strongest in a third-person view, rather than an egocentric first-person view, which suggests that this effect is not derived from visual

experience alone (Prasad & Shiffrar, 2009). Finally, a group of patients with hemiplegia (e.g., a lesion to the motor system that affects use of only one arm) were presented with point-light displays, and asked to identify the activity that was shown in the display (e.g., blowing a kiss, waving, wagging a finger, etc.). Compared to both normal and brain-damaged controls, hemiplegic patients showed deficits in identifying these actions, but *only* on the affected side (i.e., half of the presented displays were mirror-reversed) (Serino et al., 2010). Together, these results suggest that visual perception of biological motion can involve access to sensorimotor information and/or motor simulation.

Several studies have also suggested that the efficiency of perceptual processing can be affected by simultaneously executing sensorimotor movements. This has most clearly been exemplified in visual object perception: making rotational movements with one's hand can either facilitate or impair the latency in completing a mental rotation of another object (Wexler, Kosslyn, & Berthoz, 1998; Wohlschläger & Wohlschläger, 1998). This effect is sometimes quite dramatic: manually rotating an object can result in *slower* response times in mental rotation tasks at 0-degrees (i.e., requiring no mental rotation at all) than at 45-degrees (i.e., requiring a small amount of mental rotation in the same direction as manual rotation) (Wexler et al., 1998). Moreover, simply planning to make a rotational movement (but not actually executing this movement) has a similar effect on the latency to complete a mental rotation (Wohlschläger, 2001).

Perceptual identification of ambiguous or noise-embedded visual stimuli is also affected by concurrently producing movements. One set of studies has reported that perception in multistable visual displays (i.e., displays that sometimes appear as one visual pattern, and other times another) varies depending on whether experimental participants make movements that are linked to one of the stable percepts (Maruya, Yang, & Blake, 2007; Mitsumatsu, 2009; Wohlschläger, 2000). Another study examined

visual processing of shapes and letters embedded in visual noise while perceivers simultaneously wrote or drew shapes and letters (K. H. James & Gauthier, 2009). Visual identification of the letters was affected by whether or not participants concurrently drew something, but identification of shapes was not, which suggested that the greater experience that one has drawing letters (relative to shapes) made visual perception of that stimuli set more sensitive to the effects of generating movements (K. H. James & Atwood, 2009; K. H. James & Gauthier, 2006). Even more intriguingly, the identification of letters was also affected by whether the visual targets matched the drawn targets in either the straightness or curviness of written strokes, suggesting a more direct link between the motor task and visual perception. Moreover, patterns of perceptual identification showed similar patterns when participants' wrists were manually yoked to an experimenter (who was unseen, but who drew out letters and shapes during the identification task). This suggests that conscious rehearsal or high-level motor planning is not necessary in showing effects of movement on the visual identification of letters.

A last set of studies further suggests that performing actions influences perception in the auditory domain. Repp & Knoblich (2007, 2009) tested perception of tritones, which are sequential tone pairs with acoustic characteristics that make them ambiguous between sounding like a rising or falling sequence (the precise perception of rising or falling varies in the general population [Deutsch, Kuyper, & Y. Fisher, 1987]). When testing experienced pianists (versus non-piano playing musicians), perception of a tritone sequence was measured while finger sequences on a keyboard were executed. When fingering left-to-right (versus right-to-left), experienced pianists perceived more rising sequences in the tritones, while the non-piano playing musicians experienced this to a much smaller degree (Repp & Knoblich, 2007). This effect persisted when pianists merely observed an experimenter performing these actions, but only when the

orientation of the keyboard was in the same position as it would be if the perceiver were also playing it (i.e., the effect disappeared when the perceiver sat opposite the experimenter) (Repp & Knoblich, 2009).

1.3.5 Developmental accounts of action-perception couplings

Developmental research on perception-action linkages is rooted in the relatively well-known work of Gibson & Walk, who studied depth perception using the visual cliff paradigm, which tests the willingness of an infant (or an animal) to move on a glass surface under which a visible, textured surface varies in distance (E. J. Gibson & Walk, 1960; Walk, 1966). Unlike rats, which do not require visual experience in learning to avoid the “deep” side of the cliff, human infants without significant locomotor experience will crawl without hesitation over to this side. Gibson hypothesized that, at least in the case of humans, learning about depth perception was motivated by the adaptive challenges that were afforded by an infants’ typical environment (E. J. Gibson & Pick, 2000). Subsequent studies have suggested that avoiding the “deep” side of the visual cliff is correlated with crawling experience (Bertenthal, Campos, & K. C. Barrett, 1984), but this result is not always replicated (Richards & Rader, 1981, 1983). Contemporary views suggest that infants’ ability to decide what environmental conditions afford successful motor exploration develops in a piecemeal, specific, and discontinuous pattern (Adolph, 2000, 2008).

Since early work on the visual cliff, researchers have further investigated the role that various kinds of motor experience (e.g., grasping, locomoting) have on the development of perceptual representations for objects. For example, when infants manipulate objects, this often involves visual analysis of the object itself, and may lead to increased exposure and intake of visual information (Rochat, 1989). This claim is further bolstered by research showing, for example, that the ability to independently sit is

correlated with performance in a three-dimensional object completion task: this is thought to be driven by the fact that sitting is correlated with infants' experience handling and rotating objects with their hands (Soska, Adolph, & S. P. Johnson, 2010). Furthermore, experience with handling objects may also help infants learn to segregate the forms of individual objects in a visual scene (Needham, 2000). This claim is supported by research showing that young infants (without much grasping abilities) perform better in object perception tasks when given brief experience with *sticky mittens* (i.e., mittens covered in Velcro), which allow object manipulation at an earlier age than would normally be permitted (Needham, T. Barrett, & Peterman, 2002). Other research has examined, for example, the correlation of locomotor behaviour (i.e., crawling) on the developing understanding of causal movement and self-propelled objects (Cicchino & Rakison, 2008), social-emotional development (Bertenthal & Campos, 1990), and a variety of spatial abilities (see Campos et al., 2000 for review).

Exploratory motor behaviour may facilitate object perception by increasing the amount and character of infants' perceptual experiences, but what evidence is there that motor behaviour *itself* is linked to cognitive development or the understanding of actions and events? Piaget famously made the claim that motor experience is required in understanding the physical world (Piaget, 1952, 1969), but the basis and breadth of these claims have been contested by a large body of research showing sophisticated knowledge of the physical properties of objects at very early stages in infancy (Baillargeon, 1986, 1987; Baillargeon & DeVos, 1991; Baillargeon, Spelke, & Wasserman, 1985). Further work suggests that the production of actions is not always correlated with action understanding in a variety of tasks. For example, the production of means-end behaviours (e.g., pulling a board on which an interesting object is resting in order to reach it) is not always correlated with performance in perception tasks that

evaluate understanding of similar means-end behaviours (e.g., a violation-of-expectation paradigm showing videos of board-pulling) (Daum, Prinz, & Aschersleben, 2009).

Similarly, infants often perform differently when reaching *versus* looking in a variety of tasks, including A-not-B object search tasks (Baillargeon & DeVos, 1991; Baillargeon & Graber, 1988), and object-tracking tasks that involve physical barriers (Keen, 2003).

Thus, in many cases infants' motor abilities do not always echo their perceptual knowledge.

Two lines of evidence increasingly suggest, however, that there *is* a relation between the ability to perform a motor task and the understanding of actions. Providing a theoretical basis for these findings, the "Like Me" approach suggests that the basis of social cognition begins with understanding the goals or intentions of others, and that this understanding originates in applying one's own egocentric motivations and goals to the observation of others' actions (Meltzoff, 2007; Meltzoff & Brooks, 2001). Consider the goal-directedness of human actions: reaching movements are usually made towards a specific goal, or to achieve a certain end, and infants learn to encode this type of action from an early age (Carpenter, Call, & Tomasello, 2005; Woodward, 1998, 2009). Several lines of evidence further suggest that understanding of the goal-directedness of actions is an achievement in infancy that is aided by experience reaching and grasping objects oneself (i.e., facilitated with the use of *sticky mittens*) (Sommerville, Woodward, & Needham, 2005), and that the benefits from acting on objects is separable from the benefits of simply observing someone make goal-directed actions (Sommerville, Hildebrand, & Crane, 2008). This provides strong evidence that some features of goal understanding are related to the development and change of motor and reaching behaviours in infancy.

One final piece of evidence suggests that perception of visual motion recruits information about one's own motor system in infancy. This comes mainly from work on infants' perception of biological motion (Bertenthal, Proffitt, Kramer, & Spetner, 1987; Bertenthal, Proffitt, Spetner, & Thomas, 1985; Bertenthal & Campos, 1987). In one study, both 3- and 5-month-old infants were tested in displays of an adult-generated point-light display of walking or running (Booth, Pinto, & Bertenthal, 2002). Younger infants discriminated the two gait-patterns, while older infants did not, which suggests that these two groups of infants use different kinds of information to process displays of biological motion. It was hypothesized that younger infants pay more attention to low-level visual cues, like speed and joint-angles, while older infants paid more attention to high-level cues related to the motor system, like whether the limbs in these displays were symmetrically patterned. In a test of this hypothesis, point-light displays were modified such that opposing limbs (e.g., the left arm and right leg) were no longer in phase. When this change was made, 5-month-olds discriminated walking *versus* running, since the differences in symmetrical patterning were now accentuated. Indeed, further evidence hints at the possibility that this observed dependence on symmetrical patterning actually implicates the motor system more directly. While this hypothesis needs to be tested empirically, infants' ability to maintain a symmetrical phase relation between two legs improves dramatically between the ages of 3 to 6 months of age (Thelen & Ulrich, 1991). It has thus been hypothesized that 5-month-olds develop this method of perceptual analysis based on their own ability to simulate stepping when watching displays of biological motion (Bertenthal & Longo, 2008).

1.3.6 Major theoretical approaches

Research in the last 15 years has focused mostly on how generating actions can influence action perception, due in large part to the emergence of two separate

literatures. On the one hand, researchers were excited by the prospect of major breakthroughs in cognitive neuroscience based on mirror neuron research in monkey cortex (Decety et al., 1994; Gallese et al., 1996; Iacoboni et al., 1999). These findings motivated the search for representations and processes that were able to link the production of actions with perception of actions in humans. On the other hand, however, there had already been a long-standing literature on stimulus-response, or ideomotor compatibility (Greenwald, 1970; Proctor & Vu, 2006), and a literature discussing how motor control interacts with sensory feedback during the production of visually guided movements (Wolpert & Ghahramani, 2000; Wolpert, Ghahramani, & Jordan, 1995). In the contemporary study of embodied perception, two general types of accounts have emerged from these literatures, which attempt to explain some of the empirical evidence: common coding theories, and theories that appeal to forward internal models of motor control.

A common coding theory draws from William James' notions of the ideomotor principle, which suggested close ties between motor imagery and action production. Formulations of the common coding theory suggest that motor representations are coded in terms of the perceptual events that are produced by these actions (Hommel, Bertoncini, Aschersleben, & Prinz, 2001; Prinz & Hommel, 2002; Prinz, 1997): in other words, action planning and the perceptual processing of events share a common format. Consider the effects of action-effect blindness, a phenomenon on which the common coding theory was originally based (Müsseler & Hommel, 1997; Zwicker, Grosjean, & Prinz, 2010). In these paradigms, an experimental participant is given a simple motor task that is unrelated to a concurrent visual task (e.g., making right-handed button presses). Detection of a relevant visual cue (i.e., a right-pointing arrow) is impaired compared to when making other irrelevant movements (i.e., making left-handed button

presses). According to the common coding theory, this “action-effect blindness” arises because the features of events that are shared both by perception and by action planning are bound together when executing an action plan. While these features are bound together, they are less available for perceptual tasks. In other paradigms, features shared in the common format between perception and action may facilitate or enhance action production or action perception. For example, when the action plan (e.g. right-handed response) and the perceptual event (e.g., detecting a right-arrow) are functionally related (as in stimulus-response compatibility paradigms) these common formats may be primed or partially activated, and lead to facilitation of the motor response (Müsseler & Hommel, 1997).

Another approach has alternatively suggested that action-perception couplings arise from an internally generated simulation that reproduces the peripheral sensory consequences of a generated motor plan, what has come to be called “efference copy” or an “internal” or “forward” model (Wolpert et al., 1995; Wolpert & Kawato, 1998). The existence of forward models has primarily been proposed to account for the complex and dynamically varying constraints on motor control (Wolpert & Ghahramani, 2000; Wolpert, Ghahramani, & Flanagan, 2001). For example, noise inherent in the motor system, as well as quickly changing environmental constraints often results in the need to readjust a previously programmed motor movement as it is being executed. If the motor system were to wait until external sensory information about the execution of an action was processed, there would be an inherent delay integrating sensory information about the existing state of the motor system into a revised motor plan. Forward models thus provide predictive information about the sensory consequences of an action, which is integrated with actual peripheral sense information as it arrives from sense organs.

Two recent extensions of the framework defining forward models have been proposed. First, a computational implementation of this idea has been implemented: the modular selection and identification for control (MOSAIC) model (Haruno, Wolpert, & Kawato, 2001; Wolpert & Kawato, 1998). This implementation offers a computational architecture for motor control, which implements multiple forward models, and attempts to account for a variety of motor learning situations. The MOSAIC model differs from a theory of common coding, in that it assumes more stochastic and probabilistic information constructs, rather than abstract feature representations. Second, drawing broadly from notions inherent to theories invoking mental simulation, it has been hypothesized that computational structure that is used in motor control is also implicated in imitation, theory of mind, and other phenomena that involve integrating the actions of others with knowledge about one's own possible actions (Decety & Grèzes, 1999, 2006; Jeannerod, 2001; Wolpert, Doya, & Kawato, 2003). Perhaps the boldest idea is that forward models are recruited in action perception more generally, even for actions that one is not intending to produce oneself. This theoretical move attempts to generalize the properties of motor control to perceptual processes more generally, becoming, essentially, a theory of embodied perception (Wolpert et al., 2003).

1.4 Action in perception: the domain of speech

Speech perception is a field in which embodied approaches have a long history. And it seems apparent why this is so: the speech signal is different from other kinds of acoustic signals in that it is intimately tied to the movement of articulators. Speech is, in essence, the acoustic consequence of these movements. What follows below is an outline of recent empirical evidence in favour of the idea that articulatory processes are either recruited, or active in the perceptual analysis of speech. The structure of this section closely follows that of the previous one, especially with regard to the types of

empirical phenomena described. Section 1.4.1 discusses some of the historical and theoretical roots of embodied approaches in speech. Section 1.4.2 describes research that shows effects of perception-on-production, or how articulatory-motor information is active during speech perception. Section 1.4.3 describes effects of production-on-perception: how making or preparing articulatory movements can affect performance in speech perception tasks. Finally, Section 1.4.5 describes some of the relevant literature from the developmental speech perception literature that bears upon subsequent chapters.

1.4.1 Historical and theoretical roots

Motor theories of speech perception originate from early work at Haskins Laboratories, which showed that the acoustics cues used in speech are highly context-sensitive. For example, the acoustic cues to phoneme (i.e., /d/) can change dramatically in different contexts: the second formant frequency (F2) rises in the context of /di/, but falls in the context of /du/ (Liberman, Delattre, F. S. Cooper, & Gerstman, 1954). Perception studies suggest that these divergent cues (i.e., rising or falling F2) both sound like /d/ in the context of different vowels (i.e., high- or low-vowels), but sound very different when heard in isolation. In other words, a single phoneme in speech is cued by dramatically different acoustic signals in different speech contexts. And not only does a unitary phonemic percept have different acoustic instantiations, but the same acoustic cue can also signal different phonemes in different contexts as well (Liberman, Delattre, & F. S. Cooper, 1952). In this latter study, the same acoustic stop-burst was perceived as /p/ in the contexts of high-vowels /i/ and /u/, but was perceived as /k/ in the context of the low-vowel /a/, suggesting that co-articulatory information about the vowel also influences the perception of a stable acoustic cue. These findings together suggested that acoustic cues are not the primary targets of speech perception.

This work, in combination with other evidence (e.g., acoustic correlates of discrete, linguistic units are spread out in time due to the way in which our articulators prepare for upcoming gestures) suggests that speech is not an acoustic alphabet (Fowler, 1986, 1996; Goldstein & Fowler, 2003; Liberman & Mattingly, 1985). Thus, in the speech domain, two very general (but related) classes of embodied approaches have been taken: the “classical” motor theory of speech perception (Liberman, 1996; Liberman et al., 1967; Liberman & Mattingly, 1985; Liberman & Whalen, 2000), and direct-realist theories of perception (Best, 1993, 1995; Fowler, 1986, 1996). Because articulatory processes essentially cause this acoustic variability, both motor theory and direct-realist perspectives suggested that the targets of perception are articulatory gestures, rather than acoustic information.

This basic claim is supplemented by two more theoretical assertions in the Liberman & colleagues’ motor theory of speech perception: first, this theory posits that an encapsulated module dedicated to speech is recruited in perception (see Liberman & Mattingly, 1985); and second, that speech perception involves recruitment of the speech production system, since the motor system is the only way to account for the inherent variability in the acoustic signal. Because the acoustic signal is so immensely complex and appears highly context-specific, motor theorists suggested that listeners must perceive speech by using their own knowledge of the human vocal tract to re-analyze sensory input.

Direct-realist instantiations of motor theoretical approaches, despite sharing many of the general assumptions of the classical motor theory, differ on both points mentioned above. First, this theory claims that an embodied approach to the perception of speech is not special to this domain: perception of articulatory gestures is just the domain-specific instantiation of the more general kind of ecological perception previously

described in other domains, especially in vision (E. J. Gibson, 1969; J. J. Gibson, 1979). In these direct-realist approaches, the articulatory gestures of speakers are *directly* perceived by listeners in a similar way that the physical properties of objects are *directly* perceived in vision: speech perception is no more about perception of formant transitions and frequency bands (as opposed to tongue and lip gestures) than vision is about the properties of reflected light (as opposed to surfaces and shapes in the real-world). Second, this theory claims that speech perception *does* involve identification of invariant features. That is, perception is meant to identify invariant articulatory gestures by virtue of the specific (but complex) patterns that arise out of the acoustic signal (Fowler, 1996). In other words, what listeners perceive may be acoustic, but the invariant information about speech extracted from the acoustic signal is derived from gestural information.

Two newer accounts of speech perception also take an embodied approach, but assume that processes for speech production and perception are distinct (although closely linked in the brain). First, Pulvermüller (2005) suggests one such account, whereby associative links between speech perception and production are strengthened as infants begin vocalizing, and as older infants and young children begin to speak. This results in mutual activation in speech perception and production areas, and eventually forms networks of perception-action loops in cortex. Evidence for this comes from several imaging studies, where activated areas are connected in a neural network that spans superior temporal sulcus (STS), traditionally considered important in speech perception, and inferior frontal areas (Broca's area), which is traditionally considered important in speech production (Pulvermüller & Fadiga, 2010). Further work has shown, for example, that semantic information can also be included in these neural networks: simply reading words that are semantically related to different motor actions (involving,

for example, either the face, arm or leg) can selectively activate related motor areas in cortex (Hauk, Shtyrov, & Pulvermüller, 2006).

A relatively recent approach suggests that articulatory processes are recruited when the brain combines information from non-auditory (i.e., visual) and auditory modalities in speech processing (Skipper, Nusbaum, & Small, 2006; Skipper, van Wassenhove, Nusbaum, & Small, 2007; van Wassenhove, Grant, & Poeppel, 2005). Multiple streams of information are assumed in this view: visual information provides, essentially, another source of sensory input that is available for comparison to incoming auditory information, which is similar to the way that corollary sensory information is generated by internal forward models in theories of motor control (cf. Wolpert & Kawato, 1998). Critically, this internally generated information is transformed from visual to auditory information based on the speaker's own articulatory knowledge. These sensory predictions are then used to weight acoustic cues in the auditory input, and ultimately help to shape judgments about perceptual identity (see Skipper, Nusbaum & Small, 2006). While there is not yet a large body of behavioural evidence supporting this view, fMRI studies provide some support for this model. For example, recent evidence has also shown that motor activation is closely modulated by the perceptual judgements reported by participants (i.e., the perceived speech categories are correlated with articulator-specific motor activity in the brain), and that patterns of activation in sensory areas change over time, showing activation corresponding to unimodal sensory inputs at earlier stages, and fused audiovisual inputs at later stages (Skipper et al., 2007). This model, while deserving more scrutiny, provides mounting evidence that audiovisual speech activates motor areas, resulting in complex temporal patterns of integration across regions in the brain.

1.4.2 Motor activation during speech perception

Several lines of evidence have suggested that speech perception results in the activation of motor circuits related to articulation. Early behavioural work had shown, for example, that hearing speech sounds could influence the articulatory characteristics of subsequent verbal responses (W. E. Cooper, 1979). For example, hearing a repeated acoustic syllable (e.g., /pi/) while producing verbal responses has an effect on the articulatory characteristics of syllable production. Since this early work, other studies have also varied the congruency between a go/no-go auditory cue and a subsequent verbal response, similar to studies from the stimulus-response compatibility literature in vision. Spoken responses in these cases are facilitated when an auditory go-signal (i.e., a syllable) shares phonetic features with the verbal response itself (Galantucci, Fowler, & Goldstein, 2009; Gordon & Meyer, 1984; Meyer & Gordon, 1985). This is taken as evidence that the perceptual processing of the go/no-go cue can selectively facilitate certain articulatory-motor information, leading to faster, or speeded speech production. Kerzel & Bekkering (2000) reported similar effects with talking faces as the go/no-go signal, suggesting that both auditory and visual processing of a speech cue activates articulatory-motor information.

Neuroimaging (fMRI) studies have also shown that motor circuits may be involved in speech perception (Pulvermüller & Fadiga, 2010; S. K. Scott, McGettigan, & Eisner, 2009). For example, hearing speech can activate areas of primary and premotor cortex, as well as supplementary motor areas in some studies (Iacoboni & S. M. Wilson, 2006; Pulvermüller et al., 2006; S. M. Wilson, Saygin, Sereno, & Iacoboni, 2004). Similarly, perceiving audiovisual and visual speech (D. E. Callan, Jones, A. M. Callan, & Akahane-Yamada, 2004; D. E. Callan et al., 2004; Calvert & Campbell, 2003; Skipper et al., 2007) can also activate motor areas in the brain. Research from TMS studies has

also suggested that hearing speech can enhance motor-evoked potentials in speech articulators (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Watkins et al., 2003).

In summary, all of these studies are certainly suggestive of a close coupling between speech production and perception. However, they do not reveal the precise mechanisms by which these two processes are related. For example, what is the time-course of activation in these cortical areas? Are motor processes necessarily involved in the regular course of speech processing? Further research is needed to sort out these questions, but one recent report suggests that motor activation is an obligatory part of perceiving speech. In this study, the degree of contact between the tongue and palate were measured while participants produced syllables beginning with /k/ sounds (e.g., “key”). While producing these syllables, participants also listened to either congruent (i.e., the same syllable) or incongruent auditory sounds (i.e., a rhyming syllable that began instead with /t/, for example, “tea”). Even when told to produce “key” and ignore the auditory distractor (i.e., “tea”) the tongues of the participants still touched the roof of the mouth as if covertly producing /t/ sounds. These results suggest that perceiving speech results in selective, covert, and automatic activation in the speech articulators (Yuen, M. H. Davis, Brysbaert, & Rastle, 2010).

1.4.3 Sensorimotor influences in perceptual analysis of speech

In recent years, several studies have also begun testing the role that motor activation has in auditory perception. One of the first demonstrations of this phenomenon reported that the decision latencies to identify whether a word begins with a particular sound (i.e., a phoneme monitoring task) were affected by whether the participant was planning to produce a word that began with a congruent or incongruent sound (Roelofs, Özdemir, & Levelt, 2007). This evidence was taken as support for a model of speech production where speech planning and word recognition shared phonological

representations. Similarly, another recent study showed that simultaneously executing silent articulatory movements (e.g., mouth /ka/) while listening to speech (e.g., hearing acoustic /pa/) could alter the perception of speech stimuli (e.g., acoustic /pa/ was perceived as /ka/) (Sams et al., 2005). Other studies have also shown that making silent articulatory movements can influence the characteristics of evoked-potentials or neuroimaging activation in auditory cortex when hearing speech (Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; Houde, Nagarajan, Sekihara, & Merzenich, 2002; Numminen & Curio, 1999; Paus, Perry, Zatorre, Worsley, & Evans, 1996). Finally, the behavioural identification of speech sounds can be affected by repeated TMS of speech-motor cortex (D'Ausilio et al., 2009; Devlin & Watkins, 2007; Meister, S. M. Wilson, Deblieck, Wu, & Iacoboni, 2007).

The interpretation of these effects has been subject of a great deal of controversy in the literature. Some researchers have taken this evidence as support for strong versions of embodiment, claiming that the reason one finds these articulatory influences is because articulatory-motor information is necessarily recruited in the normal process of speech perception (Galantucci, Fowler, & Turvey, 2006). Others, on the other hand, have questioned whether articulatory-motor information is even the source of perceptual modulation in these studies: for example, articulation may be simply activating abstract symbolic or auditory representations in parallel, which in turn influence perception (Mahon & Caramazza, 2008). This suggests that the aforementioned research is not evidence for a central role of articulatory-motor information in speech processing, *per se*. Rather, articulatory effects may simply be a corollary to the normal auditory-only mechanisms that are used to process speech input, mediated auditory or symbolic processes that are activated in parallel with articulation (Diehl, Lotto, & Holt, 2004; Lotto

et al., 2009; S. K. Scott et al., 2009). Chapter 2 attempts to address this point more definitively.

These concerns aside, there are other gaps in this literature as well. For example, even if it were shown that articulatory-motor information was a source of perceptual influence, few studies have offered a clear explication of *what kind* of sensorimotor information triggered by articulatory processes may be important in perceptual analysis. Specifically, the mere execution of a speech movement results in a cascade of processes within the hierarchy of speech motor control: what aspects of this hierarchy might contribute to the perceptual analysis of speech? In other words, is perception selectively influenced by activation of high-level information about dynamic speech-like articulatory gestures, or can low-level information about the mere positions of the articulators (embedded even in non-speech gestures) also exert some perceptual influence? This central question about the interplay between speech perception and production is the topic of Chapter 3: precisely what kind of information is shared between articulatory-motor and perceptual processes?

1.4.4 Developmental accounts of action-perception couplings in speech

What are the ontogenetic origins of links between speech perception and production? This question remains relatively unexplored, but some evidence for the effect of perception-on-production in early speech development suggests that one's native language can influence infant's earliest vocal behaviours. For example, a mother's speech heard in the womb can shape the prosodic structure of a newborn's cry (Mampe, Friederici, Christophe, & Wermke, 2009). Prosodic qualities of vowel-like vocalizations are similarly influenced by linguistic input as early as 2 months of age (Ruzza, Rocca, Boero, & Lenti, 2006) and babbling from 6-10 months of age shows some effects of native-language exposure (de Boysson-Bardies, Sagart, & Durand,

1984; Whalen, Levitt, & Goldstein, 2007), as well as by auditory feedback more generally (i.e., deaf infants have delayed and unique patterns of babbling) (Koopmans-van Beinum, Clement, & van den Dikkenberg-Pot, 2001; Nathani, Oller, & Neal, 2007; Oller & Eilers, 1988). Finally, the production of words at older ages in infancy is also influenced by phonological and phonotactic patterns in the input, and are argued to be continuous with early babbling behaviour (de Boysson-Bardies & Vihman, 1991; McCune & Vihman, 2001; Vihman, 1991, 1993).

At the same time, some characteristics of infants' vocal behaviours, particularly babbling, often follow a more language-independent course of development (B. L. Davis & MacNeilage, 1995; MacNeilage, 1998; MacNeilage & B. L. Davis, 1990). For example, the characteristics of motor control in babbling across several languages show oscillatory behaviours that appear to stem from a more universal trait of the motor system (MacNeilage, B. L. Davis, Kinney, & Matyear, 2000). As further evidence for this point, the emergence of babbling is accompanied by similarly rhythmic movements (i.e., arm waving) at the same point in development as the onset of babbling (Iverson, Hall, Nickel, & Wozniak, 2007). Thus, there is still a significant amount of debate with regard to how developing speech abilities and early articulatory-motor behaviours are related: what is the precise influence of speech perception on early vocalization and babbling?

On the converse side, no studies have systematically examined the effect of production-on-perception, or the effect of infants' articulatory behaviours on speech perception. Although previous discussions have postulated that there may be such an influence (Kent & Vorperian, 2007; Kuhl & Meltzoff, 1988; Werker, 1993), there are several reasons for this gap in the literature. First, it is difficult to assess vocal behaviour across ages and individuals due to rapid physiological changes in infants' developing vocal tracts (Vorperian et al., 2005). Second, there are methodological difficulties in

manipulating articulatory behaviour in infant populations. Third, and most importantly, there is a long-standing theoretical argument against posing this research question: links between infant speech perception and production are not obviously predicted given their apparently asymmetrical courses of development. Infants learn to perceive sophisticated, language-specific phonetic patterns in speech before a correspondingly sophisticated system of speech production is in place (see Locke, 1983; Oller, 1980; Stark, 1980; Werker & Tees, 1999, 2005 for reviews).

Consider a recent empirical study that reports a specific asymmetry in the developmental patterns of speech perception and production (Nazzi, Bertoncini, & Bijeljic-Babic, 2009). Words produced between 12-18 months of age in many of the world's languages contain many more labial-coronal (LC) closures (e.g., *pat*) than coronal-labial (CL) closures (e.g., *tap*) (MacNeilage, B. L. Davis, Kinney, & Matyear, 1999; MacNeilage, B. L. Davis, & Matyear, 1997). However, Nazzi et al. (2009) show that a *perceptual* bias for LC words precedes this production bias: 10-month-old infants prefer to listen to LC words than CL words, even though they do not yet produce many LC words themselves. It was suggested that this perceptual bias arises from infants' sensitivity to statistical patterns in language input, rather than being linked to articulatory processes (see also Anderson, Morgan, & White, 2003; Maye, Werker, & Gerken, 2002).

On the other hand, there is other evidence to suggest that articulatory processes may still play some role in these infants' perceptual preferences. Even though an LC bias is not found in babbling, it is motivated by similar articulatory-motor constraints that *are* found in babbling (MacNeilage & B. L. Davis, 2000; MacNeilage et al., 2000). And, tellingly, 6-month-old infants, who have only just begun to make reduplicative babbles, do not show this perceptual LC bias (Nazzi et al., 2009). Without further data, however, it

is difficult to distinguish between the effects of language input from the effects of the developing articulatory-motor system.

Given this theoretical background and the corresponding lack of experimental data, it is perhaps unsurprising to find that several theoretical approaches have argued the perception-production link observed in adults does not have its roots in infancy. For example, the developmental asymmetry between perception and production has been taken as evidence against classical motor theories of speech perception, which assume that articulatory processes are recruited in speech processing (Lieberman & Mattingly, 1985). Since infants show advanced phonetic sensitivities before they have a sophisticated speech production system, the reasoning goes, then a strong interpretation of the motor theory is unlikely (Lotto et al., 2009). Others have argued that perception-production links *are* an important feature of speech processing in adults, but associative networks between speech perception and production develop only after older infants and young children have some experience producing and perceiving their own speech (Pulvermüller, 2005).

A few theoretical approaches, on the other hand, suggest that the role of articulatory knowledge in shaping infant speech perception has been underestimated. Consider a developmental account based on the direct-realist approach linking speech perception to information about gestural events. It has been suggested that gestural information is more broadly specified in early stages, but is more precisely matched to the properties of an infant's native language and the infants' own developing articulatory system in later stages of development (i.e., a differentiationist view) (Best, 1995). Yet another theoretical approach has posited the existence of mappings between articulatory, auditory, and visual information from early in infancy (Kent & Vorperian, 2007; Kuhl & Meltzoff, 1984, 1988). This is based on work showing that infants can

match cross-modally between heard vowels and talking faces: since the very first descriptions of this phenomenon, it has been suggested that cross-modal matching is made possible because of articulatory-based mappings between visual and auditory modalities when perceiving speech (Kuhl & Meltzoff, 1984, 1988; Kuhl, Williams, & Meltzoff, 1991; MacKain, Studdert-Kennedy, Spieker, & Stern, 1983; Patterson & Werker, 1999; Walton & Bower, 1993). As mentioned previously, however, direct evidence for these theoretical positions has been elusive.

In summary, several studies have suggested that there may be links between articulation and speech perception in infancy, but concrete empirical evidence to support these assertions are either mixed, or non-existent. On the one hand, some studies have suggested that early speech-like vocalizations, including babbling, are influenced by the perception of native-language speech patterns. However, other research has questioned whether speech production is driven more by its link with perception, or more by maturational factors in the motor system. On the other hand, several theorists have discussed the possibility that speech perception is tied to information about articulatory behaviours in infants, but without direct evidence implicating the articulatory-motor system.

1.5 Conclusions

The previous sections in this chapter are meant to a) define broader theoretical concepts of embodiment and differentiation, and b) review the literature relevant to action-perception linkages in the domain of vision and speech. In subsequent chapters, however, the topic of inquiry will be centred on a specific question related to the literatures reviewed above: what is the *origin* of articulatory-motor influences on speech perception, where two specific senses of “origin” are implied. First, what are the origins

of these influences in *processing* (i.e., in what sense is speech perception embodied)?

Second, what is the *developmental* origin of these influences?

This question about the origins of articulatory-motor influences *in processing* is related to several debates about links between perception and action in the domain of speech. First, critics of embodied approaches have generally argued that the link between perception and action is actually mediated by abstract or high-level conceptual representations. This, in turn suggests that the presence of action-perception couplings, per se, is not strictly evidence for embodiment, but simply indicates the presence of spreading activation between conceptual, sensory, and motor representations (Mahon & Caramazza, 2008). Merely showing dual activation of articulatory and perceptual processes is not enough to claim that perception or cognition itself is embodied: one must also understand more precisely how action and perception are related.

Second, extant theories of action-perception linkages disagree about the appropriate level of abstraction. Consider a common coding theory (Hommel et al., 2001), which is faced with the question of what levels in the motor control hierarchy (i.e., the “features” of action production) can be commonly formatted with perception (i.e., the “features” of action perception). Previous research has suggested, for example, that the commonly coded features are somewhat abstract, specifying only qualitative, rather than qualitative *and* quantitative properties of actions (Zwicker et al., 2010). Theories that appeal to forward models, on the other hand, have generally specified interactions between action and perception at relatively low levels of abstraction, which are sensitive to graded or stochastic features of actions and events (Hamilton et al., 2004). In the domain of speech, no previous work has evaluated at which levels of abstraction the links between perception and action are specified.

The first two empirical chapters address these questions about origins of articulatory-motor influences on speech perception *in processing*. The central claim in Chapter 2 suggests that sensorimotor information about the movements of articulators can indeed be a source of perceptual modulation, which is independent from symbolic or auditory representations. The central claim in Chapter 3 suggests that at least *some* action-perception linkages in speech are formatted in ways that operate below the level of linguistic representations in speech motor control. Evidence from this chapter suggests that low-level articulatory information (i.e., maintenance of a single articulatory position in the vocal tract) can influence the perception of speech, even when this information is embedded in a non-speech context.

The question of the *developmental* origins of articulatory-motor influences on speech perception is a relatively unexplored (although often considered) area of study. The basic claim made in Chapter 4 suggests that the links observed in Chapter 3 (between non-speech articulatory information and speech perception) appear early in development, and importantly before infants begin producing many speech-like vocalizations themselves. This suggests that these links do not develop from experience speaking, *per se*, but rather from early, experience-independent mappings between the articulators and audiovisual speech information.

Chapter 5, the final chapter of this dissertation will present some implications of these findings for language researchers, and for psychological understanding of cognitive processes more broadly. The central objectives in this dissertation will be reinforced, and a summary of the evidence from these studies will be provided. It is hoped that this research will offer insight into how perceptual analysis in the domain of speech can be tied to the motor system.

1.6 References

- Adolph, K. E. (2000). Specificity of learning: Why infants fall over a veritable cliff. *Psychological Science*, 11(4), 290-295.
- Adolph, K. E. (2008). Learning to move. *Current Directions in Psychological Science*, 17(3), 213-218.
- Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, 46(2-3), 155-182.
- Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28(2), 105-167. doi:10.1017/S0140525X05000038
- Aslin, R. N., Jusczyk, P. W., & Pisoni, D. B. (1998). Speech and auditory processing during infancy: Constraints on and precursors to language. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology* (Vol. 2, pp. 147–254). New York: Wiley.
- Baillargeon, R. (1986). Representing the existence and the location of hidden objects: Object permanence in 6-and 8-month-old infants. *Cognition*, 23(1), 21–41.
- Baillargeon, R. (1987). Object permanence in 3½-and 4½-month-old infants. *Developmental Psychology*, 23(5), 655–664.
- Baillargeon, R., & DeVos, J. (1991). Object permanence in young infants: Further evidence. *Child Development*, 62(6), 1227–1246.
- Baillargeon, R., & Graber, M. (1988). Evidence of location memory in 8-month-old infants in a nonsearch AB task. *Developmental Psychology*, 24(4), 502–511.
- Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in 5.5-month-old infants. *Cognition*, 20, 191–208.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4), 723–742.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.
- Berkeley, G. (1709). *An essay towards a new theory of vision*. New York: Dutton.
- Bertenthal, B. I., & Campos, J. J. (1990). A systems approach to the organizing effects of self-produced locomotion during infancy. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 6, pp. 1–60). Norwood, NJ: Ablex.
- Bertenthal, B. I., & Campos, J. J. (1987). New directions in the study of early experience. *Child Development*, 58(3), 560-567. doi:10.2307/1130198
- Bertenthal, B. I., Campos, J. J., & Barrett, K. C. (1984). Self-produced locomotion: An organizer of emotional, cognitive, and social development in infancy. In R. N. Emde & R. Harmon (Eds.), *Continuities and discontinuities in development* (pp. 175–210). New York, NY: Plenum Press.

- Bertenthal, B. I., & Longo, M. R. (2008). Motor knowledge and action understanding: A developmental perspective. In R. L. Klatzky, B. MacWhinney, & M. Behrman (Eds.), *Embodiment, ego-space, and action*, Carnegie Mellon Symposia on Cognition (pp. 323-368). New York: Psychology Press.
- Bertenthal, B. I., Proffitt, D. R., Kramer, S. J., & Spetner, N. B. (1987). Infants' encoding of kinetic displays varying in relative coherence. *Developmental Psychology*, 23(2), 171-178. doi:10.1037/0012-1649.23.2.171
- Bertenthal, B. I., Proffitt, D. R., Spetner, N. B., & Thomas, M. A. (1985). The development of infant sensitivity to biomechanical motions. *Child Development*, 56(3), 531-543. doi:10.2307/1129742
- Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. F. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life*, NATO Science Series (Vol. 69, pp. 289-304). Norwell, MA: Kluwer.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange & J. J. Jenkins (Eds.), *Speech perception and linguistic experience: Issues in cross-language research*, Cross-language speech perception (pp. 171-204). Timonium, MD: York Press.
- Bingham, G. P. (1987). Kinematic form and scaling: Further investigations on the visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance*, 13(2), 155-177.
- Bloom, P. (1996). Intention, history, and artifact concepts. *Cognition*, 60(1), 1-29.
- Booth, A. E., Pinto, J., & Bertenthal, B. I. (2002). Perception of the symmetrical patterning of human gait by infants. *Developmental Psychology*, 38(4), 554-563. doi:10.1037/0012-1649.38.4.554
- Bosbach, S., Cole, J., Prinz, W., & Knoblich, G. (2005). Inferring another's expectation from action: the role of peripheral sensation. *Nature Neuroscience*, 8(10), 1295-1297.
- de Boysson-Bardies, B., Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language*, 11(1), 1-15.
- de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67(2), 297-319.
- Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, 106(1-2), 3-22.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47(1-3), 139-159.

- Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage*, 22(3), 1182-1194. doi:10.1016/j.neuroimage.2004.03.006
- Callan, D. E., Jones, J. A., Munhall, K. G., Kroos, C., Callan, A. M., & Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *Journal of Cognitive Neuroscience*, 16(5), 805-816. doi:10.1162/089892904970771
- Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *Journal of Cognitive Neuroscience*, 15(1), 57-70. doi:10.1162/089892903321107828
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, 15(8), 1243-1249. doi:10.1093/cercor/bhi007
- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, 16(19), 1905-1910. doi:10.1016/j.cub.2006.07.065
- Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., & Witherington, D. (2000). Travel broadens the mind. *Infancy*, 1(2), 149. doi:10.1207/S15327078IN0102_1
- Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science*, 8(1), F13-F20.
- Chao, L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage*, 12(4), 478-484.
- Chiel, H. J., & Beer, R. D. (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences*, 20(12), 553-557.
- Cicchino, J. B., & Rakison, D. H. (2008). Producing and processing self-propelled motion in infancy. *Developmental Psychology*, 44(5), 1232-1241. doi:10.1037/a0012619
- Clark, A. (1997). *Being there: Putting brain, body, and world together*. Cambridge, MA: MIT Press.
- Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, 3(9), 345-351.
- Cohen, L., Chaput, H., & Cashion, C. (2002). A constructivist model of infant cognition. *Cognitive Development*, 17(3-4), 1323-1343.
- Cooper, W. E. (1979). *Speech production and perception: Studies in selective adaptation*. Norwood, NJ: Ablex.
- Craighero, L., Fadiga, L., Rizzolatti, G., & Umiltà, C. (1999). Action for perception: A motor-visual attentional effect. *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1673-1692.

- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19(5), 381-385. doi:10.1016/j.cub.2009.01.017
- Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2005). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, 9, 28-30.
- Daum, M. M., Prinz, W., & Aschersleben, G. (2009). Means-end behavior in young infants: The interplay of action perception and action production. *Infancy*, 14(6), 613-640. doi:10.1080/15250000903263965
- Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech & Hearing Research*, 38(6), 1199-1211.
- Decety, J., & Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences*, 3(5), 172-178.
- Decety, J., & Grèzes, J. (2006). The power of simulation: Imagining one's own and other's behavior. *Brain Research*, 1079(1), 4-14.
- Decety, J., & Jeannerod, M. (1995). Mentally simulated movements in virtual reality: does Fitt's law hold in motor imagery? *Behavioural Brain Research*, 72(1-2), 127-134. doi:10.1016/0166-4328(96)00141-6
- Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., Mazziotta, J. C., et al. (1994). Mapping motor representations with positron emission tomography. *Nature*, 371(6498), 600-602. doi:10.1038/371600a0
- Dennett, D. C. (1986). *Content and Consciousness*. London: Routledge.
- Deutsch, D., Kuyper, W. L., & Fisher, Y. (1987). The tritone paradox: Its presence and form of distribution in a general population. *Music Perception*, 5(1), 79-92.
- Devlin, J. T., & Watkins, K. E. (2007). Stimulating language: Insights from TMS. *Brain: A Journal of Neurology*, 130(3), 610-622. doi:10.1093/brain/awl331
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149-179.
- Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology*, 19(6), 643-647. doi:10.1111/j.1469-8986.1982.tb02516.x
- Edwards, M. G., Humphreys, G. W., & Castiello, U. (2003). Motor facilitation following action observation: A behavioural study in prehensile action. *Brain and Cognition*, 53(3), 495-502.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303.
- Elman, J., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15(2), 399-402.
- Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, 17(8), 1703-1714. doi:10.1046/j.1460-9568.2003.02601.x
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Fogassi, L., & Ferrari, P. F. (2007). Mirror neurons and the evolution of embodied language. *Current Directions in Psychological Science*, 16(3), 136-141. doi:10.1111/j.1467-8721.2007.00491.x
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14(1), 3-28.
- Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, 99(3), 1730-41. doi:8819862
- Fowler, C. A., & Galantucci, B. (2005). The relation of speech perception and speech production. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 633-652). Hoboken, NJ: Wiley.
- Funk, M., Shiffrar, M., & Brugger, P. (2005). Hand movement observation by individuals born without hands: Phantom limb experience constrains visual limb perception. *Experimental Brain Research*, 164(3), 341-346.
- Galantucci, B., Fowler, C. A., & Goldstein, L. (2009). Perceptuomotor compatibility effects in speech. *Attention, Perception & Psychophysics*, 71(5), 1138-1149. doi:10.3758/APP.71.5.1138
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361-377.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain: A Journal of Neurology*, 119 (Pt 2), 593-609. doi:8800951
- Gibbs, R. W. (2006). *Embodiment and cognitive science*. New York: Cambridge University Press.
- Gibson, E. J., & Walk, R. D. (1960). The visual cliff. *Scientific American*, 202(4), 64-71.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. East Norwalk, CT: Appleton-Century-Crofts.
- Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development*. New York: Oxford University Press.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

- Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning; differentiation or enrichment? *Psychological Review*, 62(1), 32-41.
- Glenberg, A. M. (2008). Toward the integration of bodily states, language, and action. In G. R. Semin & E. R. Smith (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*. (pp. 43-70). New York: Cambridge University Press.
- Glenberg, A. M., & Kaschak, M. P. (2003). The body's contribution to language. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 43. (pp. 93-126). New York: Elsevier Science.
- Goldman, A. (2006). *Simulating minds*. New York: Oxford University Press.
- Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities*, Phonology and Phonetics (pp. 159-207). New York: Mouton.
- Gordon, P. C., & Meyer, D. E. (1984). Perceptual-motor processing of phonetic features in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 10(2), 153-178.
- Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideomotor mechanism. *Psychological Review*, 77(2), 73-99. doi:10.1037/h0028689
- Grèzes, J., & Decety, J. (2002). Does visual perception of object afford action? Evidence from a neuroimaging study. *Neuropsychologia*, 40(2), 212-222.
- Grosjean, M., Shiffrar, M., & Knoblich, G. (2007). Fitts's Law Holds for Action Perception. *Psychological Science*, 18(2), 95-99. doi:10.1111/j.1467-9280.2007.01854.x
- Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27(03), 377-396.
- Hamilton, A., Wolpert, D., & Frith, U. (2004). Your own action influences how you perceive another person's action. *Current Biology*, 14(6), 493-498. doi:10.1016/j.cub.2004.03.007
- Handy, T. C., Borg, J. S., Turk, D. J., Tipper, C. M., Grafton, S. T., & Gazzaniga, M. S. (2005). Placing a tool in the spotlight: Spatial attention modulates visuomotor responses in cortex. *NeuroImage*, 26(1), 266-276.
- Handy, T. C., Grafton, S. T., Shroff, N. M., Ketay, S., & Gazzaniga, M. S. (2003). Graspable objects grab attention when the potential for action is recognized. *Nature Neuroscience*, 6(4), 421-427.
- Handy, T. C., Tipper, C. M., Borg, J. S., Grafton, S. T., & Gazzaniga, M. S. (2006). Motor experience with graspable objects reduces their implicit analysis in visual-and motor-related cortex. *Brain Research*, 1097(1), 156-166.
- Haruno, M., Wolpert, D. M., & Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Computation*, 13(10), 2201-2220.

- Haslinger, B., Erhard, P., Altenmüller, E., Hennenlotter, A., Schwaiger, M., Einsiedel, H. G. V., Rummeny, E., et al. (2004). Reduced recruitment of motor association areas during bimanual coordination in concert pianists. *Human Brain Mapping*, 22(3), 206-215. doi:10.1002/hbm.20028
- Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., & Ceballos-Baumann, A. O. (2005). Transmodal sensorimotor networks during action observation in professional pianists. *Journal of Cognitive Neuroscience*, 17(2), 282-293.
- Haueisen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, 13(6), 786-792.
- Hauk, O., Shtyrov, Y., & Pulvermüller, F. (2006). The sound of actions as reflected by mismatch negativity: Rapid activation of cortical sensory & motor networks by sounds associated with finger and tongue movements. *European Journal of Neuroscience*, 23(3), 811-821. doi:10.1111/j.1460-9568.2006.04586.x
- Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, 42(2), 180-190. doi:10.1111/j.1469-8986.2005.00272.x
- Helmholtz, H. V. (1867). *A treatise on physiological optics*. (J. P. Southall, Tran.). New York: Dover.
- Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology*, 40(2), 129-141. doi:10.1016/S0167-8760(00)00161-6
- Hommel, B., Bertoncini, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24(05), 849-878.
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, 14(8), 1125-1138. doi:10.1162/089892902760807140
- Iacoboni, M., & Wilson, S. M. (2006). Beyond a single area: Motor control and language within a neural architecture encompassing Broca's area. *Cortex*, 42(4), 503.
- Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, 286(5449), 2526-2528. doi:10.1126/science.286.5449.2526
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1245-1248. doi:10.1073/pnas.0810063106
- Iverson, J. M., Hall, A. J., Nickel, L., & Wozniak, R. H. (2007). The relationship between reduplicated babble onset and laterality biases in infant rhythmic arm movements. *Brain and Language*, 101(3), 198-207. doi:10.1016/j.bandl.2006.11.004

- James, K. H., & Atwood, T. P. (2009). The role of sensorimotor learning in the perception of letter-like forms: Tracking the causes of neural specialization for letters. *Cognitive Neuropsychology*, 26(1), 91-110. doi:10.1080/02643290802425914
- James, K. H., & Gauthier, I. (2006). Letter processing automatically recruits a sensory-motor brain network. *Neuropsychologia*, 44(14), 2937-2949. doi:10.1016/j.neuropsychologia.2006.06.026
- James, K. H., & Gauthier, I. (2009). When writing impairs reading: Letter perception's susceptibility to motor interference. *Journal of Experimental Psychology: General*, 138(3), 416-431. doi:10.1037/a0015836
- James, W. (1890). *The principles of psychology*. New York: Holt.
- Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *NeuroImage*, 14(1), S103-S109.
- Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: The cortical mechanisms of visuomotor transformation. *Trends in Neurosciences*, 18(7), 314-320.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2), 201-211.
- Keen, R. (2003). Representation of objects and events: Why do infants look so smart and toddlers look so dumb? *Current Directions in Psychological Science*, 12(3), 79-83. doi:10.1111/1467-8721.01234
- Kelso, J. A. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: The MIT Press.
- Kent, R. D., & Vorperian, H. K. (2007). In the mouths of babes: Anatomic, motor, and sensory foundations of speech development in children. In R. Paul (Ed.), *Language disorders from a developmental perspective: Essays in honor of Robin S. Chapman*. (pp. 55-81). Mahwah, NJ: Lawrence Erlbaum.
- Kerzel, D., & Bekkering, H. (2000). Motor activation from visible speech: Evidence from stimulus response compatibility. *Journal of Experimental Psychology: Human Perception and Performance*, 26(2), 634-647. doi:10.1037/0096-1523.26.2.634
- Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4), 513-549.
- Knoblich, G. (2008). Motor contributions to action perception. In R. Klatzky, B. MacWhinney, & M. Berhmann (Eds.), *Embodiment, Ego-Space, and Action* (pp. 45-78). New York: Psychology Press.
- Knoblich, G., & Flach, R. (2001). Predicting the Effects of Actions: Interactions of Perception and Action. *Psychological Science*, 12(6), 467-472. doi:10.1111/1467-9280.00387

- Knoblich, G., Seigerschmidt, E., Flach, R., & Prinz, W. (2002). Authorship effects in the prediction of handwriting strokes: Evidence for action simulation during action perception. *The Quarterly Journal of Experimental Psychology Section A*, 55(3), 1027–1046.
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297(5582), 846-848.
- Koopmans-van Beinum, F. J., Clement, C. J., & van den Dikkenberg-Pot, I. (2001). Babbling and the lack of auditory speech perception: A matter of coordination? *Developmental Science*, 4(1), 61-70. doi:10.1111/1467-7687.00149
- Korte, A. (1915). Kinematoskopische untersuchungen [Kinematoscopic investigations]. *Zeitschrift für Psychologie*, 72, 194-296.
- Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, NJ: Erlbaum.
- Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior & Development*, 7(3), 361-381.
- Kuhl, P. K., & Meltzoff, A. N. (1988). Speech as an intermodal object of perception. In A. Yonas (Ed.), *Perceptual development in infancy*, The Minnesota Symposium on Child Development. (Vol. 20, pp. 235-266). Hillsdale, NJ: Lawrence Erlbaum.
- Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15), 9096-9101.
- 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.
- Lacquanti, F., Terzuolo, C., & Viviani, P. (1983). The law relating kinematic and figural aspects of drawing movements. *Acta Psychologica*, 54, 115-130.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Lepage, J. F., & Théret, H. (2006). EEG evidence for the presence of an action observation-execution matching system in children. *European Journal of Neuroscience*, 23(9), 2505-2510.
- Lepage, J. F., & Théret, H. (2007). The mirror neuron system: Grasping others' actions from birth? *Developmental Science*, 10(5), 513-523. doi:10.1111/j.1467-7687.2007.00631.x
- Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*.
- Lhermitte, F., Pillon, B., & Serdaru, M. (1986). Human autonomy and the frontal lobes. Part I: Imitation and utilization behavior: a neuropsychological study of 75 patients. *Annals of Neurology*, 19(4), 326-34.

- Liberman, A. M. (1996). *Speech: A Special Code*. MIT Press.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431-61.
- Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, 65, 497-516. doi:10.2307/1418032
- Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, 68(8), 13.
- 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
- Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4(5), 187 - 196. doi:10.1016/S1364-6613(00)01471-6
- Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Publishers.
- Lotto, A. J., Hickok, G., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13(3), 110-114. doi:10.1016/j.tics.2008.11.008
- Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing People From Their Movement. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1), 210-220. doi:10.1037/0096-1523.31.1.210
- MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science*, 219(4590), 1347-1349. doi:10.1126/science.6828865
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21(4), 499-546. doi:10.1017/S0140525X98001265
- MacNeilage, P. F., & Davis, B. L. (1990). Acquisition of speech production: Frames, then content. (M. Jeannerod, Ed.) *Attention and performance 13: Motor representation and control*, 453-476.
- MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, 288(5465), 527-531. doi:10.1126/science.288.5465.527
- MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (1999). Origin of serial-output complexity in speech. *Psychological Science*, 459-460.
- MacNeilage, P. F., Davis, B. L., Kinney, A., & Matyear, C. L. (2000). The motor core of speech: A comparison of serial organization patterns in infants and languages. *Child Development*, 71(1), 153-163. doi:10.1111/1467-8624.00129
- MacNeilage, P. F., Davis, B. L., & Matyear, C. L. (1997). Babbling and first words: Phonetic similarities and differences. *Speech Communication*, 22(2-3), 269-277.

- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, 102(1-3), 59-70. doi:10.1016/j.jphysparis.2008.03.004
- 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.2009.09.064
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Holt.
- Maruya, K., Yang, E., & Blake, R. (2007). Voluntary action influences visual competition. *Psychological Science*, 18(12), 1090-1098. doi:10.1111/j.1467-9280.2007.02030.x
- Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101-B111.
- McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research*, 44(3), 670-684. doi:10.1044/1092-4388(2001/054)
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17(19), 1692-1696. doi:10.1016/j.cub.2007.08.064
- Meltzoff, A. N. (2007). The 'like me' framework for recognizing and becoming an intentional agent. *Acta Psychologica*, 124(1), 26-43.
- Meltzoff, A. N., & Brooks, R. (2001). "Like me" as a building block for understanding other minds: Bodily acts, attention, and intention. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 171-191). Cambridge, MA: MIT Press.
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198(4312), 75-78. doi:10.1126/science.198.4312.75
- Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, 54(3), 702-709. doi:10.2307/1130058
- Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development & Parenting, Perceptual Development*, 6(3), 179-192.
- Merleau-Ponty, M. (1962). *Phenomenology of perception*. (C. Smith, Tran.). London: Routledge.
- Meyer, D. E., & Gordon, P. C. (1985). Speech production - motor programming of phonetic features. *Journal of Memory and Language*, 24(1), 3-26.
- Mitsumatsu, H. (2009). Voluntary action affects perception of bistable motion display. *Perception*, 38(10), 1522-1535. doi:10.1068/p6298

- Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, 29(31), 9819-9825. doi:10.1523/JNEUROSCI.6018-08.2009
- Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review*, 104, 686-713.
- Müsseler, J., & Hommel, B. (1997). Blindness to response-compatible stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 861-872. doi:10.1037/0096-1523.23.3.861
- Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental Science*, 13(3), 407-420. doi:10.1111/j.1467-7687.2009.00898.x
- Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), 20470-20475.
- Nathani, S., Oller, D. K., & Neal, A. R. (2007). On the robustness of vocal development: An examination of infants with moderate-to-severe hearing loss and additional risk factors. *Journal of Speech, Language, and Hearing Research*, 50(6), 1425-1444. doi:10.1044/1092-4388(2007/099)
- Nazzi, T., Bertoncini, J., & Bijeljac-Babic, R. (2009). A perceptual equivalent of the labial-coronal effect in the first year of life. *The Journal of the Acoustical Society of America*, 126, 1440-1446.
- Needham, A. (2000). Improvements in object exploration skills may facilitate the development of object segregation in early infancy. *Journal of Cognition and Development*, 1(2), 131-156.
- Needham, A., Barrett, T., & Peterman, K. (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using 'sticky mittens' enhances young infants' object exploration skills. *Infant Behavior and Development*, 25(3), 279-295. doi:10.1016/S0163-6383(02)00097-8
- Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters*, 272(1), 29-32.
- Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., & Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research*, 24(2), 190-198.
- Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komishan, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology, Vol. 1: Production* (pp. 92-112). New York: Academic Press.
- Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, 59(2), 441-449. doi:3359864

- Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior & Development*, 22(2), 237-247.
- Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., & Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: Role of motor-to-sensory discharges. *European Journal of Neuroscience*, 8(11), 2236-2246.
- Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.
- Piaget, J. (1969). *The mechanisms of perception*. New York: Basic Books.
- Plamondon, R., & Alimi, A. M. (1997). Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences*, 20(02), 279-303.
- Prasad, S., & Shiffrar, M. (2009). Viewpoint and the recognition of people from their movements. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 39-49. doi:10.1037/a0012728
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9(2), 129-154. doi:10.1080/713752551
- Prinz, W., & Hommel, B. (2002). *Common mechanisms in perception and action*. New York: Oxford University Press.
- Proctor, R. W., & Vu, K. L. (2006). *Stimulus-Response Compatibility Principles: Data, Theory, and Application*. Boca Raton, FL: CRC Press.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576-582. doi:10.1038/nrn1706
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351-360. doi:10.1038/nrn2811
- Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., 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
- Repp, B. H., & Knoblich, G. (2007). Action can affect auditory perception. *Psychological Science*, 18(1), 6-7.
- Repp, B. H. (1981). On levels of description in speech research. *Journal of the Acoustical Society of America*, 69(5), 1462-1464.
- Repp, B. H., & Knoblich, G. (2009). Performed or observed keyboard actions affect pianists' judgements of relative pitch. *The Quarterly Journal of Experimental Psychology*, 62(11), 2156-2170. doi:10.1080/17470210902745009
- Richards, J. E., & Rader, N. (1981). Crawling-onset age predicts visual cliff avoidance in infants. *Journal of Experimental Psychology: Human Perception and Performance*, 7(2), 382-387.

- Richards, J. E., & Rader, N. (1983). Affective, behavioral, and avoidance responses on the visual cliff: Effects of crawling onset age, crawling experience, and testing age. *Psychophysiology*, 20(6), 633–642.
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5), 188-194. doi:10.1016/S0166-2236(98)01260-0
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169-92. doi:15217330
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3(2), 131-141.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews. Neuroscience*, 2(9), 661-70. doi:11533734
- Rochat, P. (1989). Object manipulation and exploration in 2-to 5-month-old infants. *Developmental Psychology*, 25(6), 871–884.
- Roelofs, A., Özdemir, R., & Levelt, W. J. M. (2007). Influences of spoken word planning on speech recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 900-913. doi:10.1037/0278-7393.33.5.900
- Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance*, 7(4), 733–740.
- Ruzza, B., Rocca, F., Boero, D. L., & Lenti, C. (2006). Investigating the musical qualities of early infant sounds. *Annals of the New York Academy of Sciences*, 999, 527–529.
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, 23(2-3), 429-435.
- Sato, W., & Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition*, 104(1), 1-18. doi:S0010-0277(06)00120-X
- Sawusch, J. R., & Gagnon, D. A. (1995). Auditory coding, cues, and coherence in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3), 635-652. doi:10.1037/0096-1523.21.3.635
- Scheerer, E. (1987). Muscle sense and innervation feelings: A chapter in the history of perception and action. In H. Heuer & A. F. Sanders (Eds.), *Perspectives in perception and action* (pp. 171-194). Hillsdale, NJ: Erlbaum.
- Schütz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: action-induced modulation of perception. *Trends in Cognitive Sciences*, 11(8), 349-355. doi:10.1016/j.tics.2007.06.005
- Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, 16(4), 197-201.

- Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action - candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, 10(4), 295-302. doi:10.1038/nrn2603
- Serino, A., De Filippo, L., Casavecchia, C., Coccia, M., Shiffrar, M., & Làdavas, E. (2010). Lesions to the motor system affect action perception. *Journal of Cognitive Neuroscience*, 22(3), 413-426. doi:10.1162/jocn.2009.21206
- Shiffrar, M., & Freyd, J. J. (1990). Apparent motion of the human body. *Psychological Science*, 1(4), 257-264.
- Shiffrar, M., & Freyd, J. J. (1993). Timing and apparent motion path choice with human body photographs. *Psychological Science*, 379-384.
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2006). Lending a helping hand to hearing: another motor theory of speech perception. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 250 - 285). Cambridge, UK: Cambridge University Press.
- Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17(10), 2387-2399. doi:10.1093/cercor/bhl147
- Sommerville, J. A., Hildebrand, E. A., & Crane, C. C. (2008). Experience matters: The impact of doing versus watching on infants' subsequent perception of tool-use events. *Developmental Psychology*, 44(5), 1249-1256. doi:10.1037/a0012296
- Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, 96(1), B1-B11. doi:10.1016/j.cognition.2004.07.004
- Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates 3D object completion. *Developmental Psychology*, 46(1), 129-138.
- Spelke, E. (1995). Initial knowledge: six suggestions. *Cognition*, 433-447.
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1), 89-96.
- Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komishan, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol 1., Production* (Vol. 1, pp. 73-92). New York: Academic Press.
- Stevens, J. A., Fonlupt, P., Shiffrar, M., & Decety, J. (2000). New aspects of motion perception: selective neural encoding of apparent human movements. *NeuroReport*, 11(1), 109-115.
- Thelen, E., & Fisher, D. M. (1982). Newborn stepping: An explanation for a "disappearing" reflex. *Developmental Psychology*, 18(5), 760-775. doi:10.1037/0012-1649.18.5.760
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

- Thelen, E., & Ulrich, B. D. (1991). Hidden skills: A dynamic systems analysis of treadmill stepping during the first year. *Monographs of the Society for Research in Child Development*, 56(1), 1–103.
- Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry*, 42(1), 3–48.
- Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 830–846.
- Vihman, M. M. (1991). Ontogeny of phonetic gestures: Speech production. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception: Proceedings of a conference to honor Alvin M. Liberman*. Mahwah, NJ: Lawrence Erlbaum.
- Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21(1/2), 61–82.
- Viviani, P. (2002). Motor competence in the perception of dynamic events: a tutorial. In W. Prinz & B. Hommel (Eds.), *Common Mechanisms in Perception and Action*, Attention and Performance (pp. 406–442). New York: Oxford University Press.
- Viviani, P., & Schneider, R. (1991). A developmental study of the relationship between geometry and kinematics in drawing movements. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 198–218.
- Viviani, P., & Stucchi, N. (1989). The effect of movement velocity on form perception: Geometric illusions in dynamic displays. *Perception and Psychophysics*, 46(3), 266–274.
- Viviani, P., & Stucchi, N. (1992). Biological movements look constant: Evidence if motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 198–218.
- Viviani, P., & Terzuolo, C. (1982). Trajectory determines movement dynamics. *Neuroscience*, 7, 431–437.
- Vogt, S., Taylor, P., & Hopkins, B. (2003). Visuomotor priming by pictures of hand postures: perspective matters. *Neuropsychologia*, 41(8), 941–951.
- Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, 117, 338–350.
- Walk, R. D. (1966). The development of depth perception in animals and human infants. *Monographs of the Society for Research in Child Development*, 82–108.
- Walton, G. E., & Bower, T. G. R. (1993). Amodal representation of speech in infants. *Infant Behavior and Development*, 16(2), 233–243.
- van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, 102(4), 1181–1186.

- Watkins, K. E., Strafella, A. 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
- Werker, J. F. (1993). The contribution of the relation between vocal production and perception to a developing phonological system. *Journal of Phonetics*, Phonetic development, 21(1-2), 177-180.
- Werker, J. F. (1995). Exploring developmental changes in cross-language speech perception. In L. R. Gleitman & M. Liberman (Eds.), *Language: An invitation to cognitive science*, Vol. 1 (2nd ed.). An invitation to cognitive science (pp. 87-106). Cambridge, MA: Mit Press.
- 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.
- Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, 50, 509-535.
- Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, 46(3), 233-251. doi:10.1002/dev.20060
- Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, 68(1), 77-94.
- Whalen, D. H., Levitt, A. G., & Goldstein, L. M. (2007). VOT in the babbling of French- and English-learning infants. *Journal of Phonetics*, 35(3), 341-352. doi:10.1016/j.wocn.2006.10.001
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625-636.
- 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), 701-702. doi:10.1038/nn1263
- Wohlschläger, A. (2000). Visual motion priming by invisible actions. *Vision Research*, 40(8), 925-930. doi:10.1016/S0042-6989(99)00239-4
- Wohlschläger, A. (2001). Mental object rotation and the planning of hand movements. *Perception & Psychophysics*, 63(4), 709-718.
- Wohlschläger, A., & Bekkering, H. (2002). Is human imitation based on a mirror-neurone system? Some behavioural evidence. *Experimental Brain Research*, 143(3), 335-341.
- Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 397-412. doi:10.1037/0096-1523.24.2.397

- Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1431), 593-602. doi:10.1098/rstb.2002.1238
- Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, 3, 1212-1217.
- Wolpert, D. M., Ghahramani, Z., & Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences*, 5(11), 487 - 494. doi:10.1016/S1364-6613(00)01773-3
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269(5232), 1880.
- Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7-8), 1317-1329. doi:10.1016/S0893-6080(98)00066-5
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1), 1-34.
- Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current Directions in Psychological Science*, 18(1), 53-57. doi:10.1111/j.1467-8721.2009.01605.x
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-Month-olds use distinct objects as cues to categorize speech information. *Cognition*, 113(2), 234-243. doi:10.1016/j.cognition.2009.08.010
- Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*. doi:10.1111/j.1532-7078.2009.00024.x
- Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 107(2), 592-597. doi:10.1073/pnas.0904774107
- Zwaan, R. A., & Taylor, L. J. (2006). Seeing, Acting, Understanding: Motor Resonance in Language Comprehension. *Journal of Experimental Psychology: General*, 135(1), 1-11.
- Zwicker, J., Grosjean, M., & Prinz, W. (2010). On interference effects in concurrent perception and action. *Psychological Research*, 74(2), 152-171. doi:10.1007/s00426-009-0226-2

2: SENSORIMOTOR ASPECTS OF ARTICULATION MODULATE AUDITORY SPEECH PERCEPTION ¹

Embodied approaches to perception suggest that some types of perceptual analyses are based on motor simulation or information from the motor system (e.g., Haruno, Wolpert, & Kawato, 2001; Jeannerod, 2001; Prinz & Hommel, 2002; Rizzolatti & Craighero, 2004). This idea has a long history in speech research, where motor theories of speech perception have suggested that speech processing involves an innate and specialized articulatory (i.e., speech-producing) module (Liberman & Mattingly, 1985), or recovers articulatory gestures through domain-general mechanisms (Fowler, 1986; Best, 1995). Classical theories in speech research focus instead on purely auditory strategies in perceptual processing, advocating against a specific role for articulatory-motor information (Diehl, Lotto, & Holt, 2004). Current debates about the nature of articulatory influences in speech perception illustrate how controversial this topic remains (Lotto, Hickok, & Holt, 2009; Massaro & Chen, 2008; Pulvermüller & Fadiga, 2010; Scott, McGettigan, & Eisner, 2009).

One type of evidence that implicates the motor system in perceptual analysis of speech comes from demonstrations of motor activation during speech perception. Perceiving auditory speech activates motor areas in the brain, specifically areas that control the articulators used to produce the specific phonemes that are heard (Wilson, Saygin, Sereno, & Iacoboni, 2004; Pulvermüller et al., 2006). Similarly, perceiving speech while receiving trans-cranial magnetic stimulation (TMS) selectively enhances

¹ A version of this chapter has been submitted for publication. Yeung, H. H., Scott, M., Gick, B., & Werker, J. F. (2010) Sensorimotor Aspects of Articulation Modulate Auditory Speech Perception.

motor-evoked potentials in speech articulators (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Watkins, Strafella, & Paus, 2003). Moreover, as shown in behavioural studies, hearing speech that is incongruent with a planned utterance can delay, and sometimes even alter the quality of speech articulation (Galantucci, Fowler, & Goldstein, 2009; Gordon & Meyer, 1984; Houde & Jordan, 1998). Recent evidence has further reported that hearing speech involves automatic and involuntary activation of compatible articulatory gestures (Yuen, Davis, Brysbaert, & Rastle, 2010).

Together, this evidence provides support for the idea that speech perception and production are closely linked, and that perceptual information is readily available for guiding and activating articulatory processes. Classical models, however, make different assumptions about encapsulation within speech production (Levelt, 2001) versus perception (Diehl et al., 2004). Thus, reported effects of perception-on-production are far less controversial than are claims of production-on-perception.

An early example of this second type of pattern, which shows speech production influencing perception, comes from work demonstrating that seeing (McGurk & MacDonald, 1976) or touching (Fowler & Dekle, 1991) a talking face can alter auditory speech perception. It has been assumed that this information about talking faces in visual or haptic modalities may be directly linked to articulatory processes, which play a critical role in the perceptual processing of speech (Fowler & Galantucci, 2005; Skipper, Nusbaum, & Small, 2006). This articulatory claim is controversial, however. Rather than illustrating articulatory-motor influences in speech perception, visual or haptic speech could alternatively activate abstract, perhaps phonemic representations. Subsequent activation of auditory and/or motor processes might originate from top-down feedback from these abstract (e.g., phonological) levels of processing, activated independently by

sensory information in visual or tactile/haptic modalities (Massaro & Chen, 2008; Mahon & Caramazza, 2008).

Recent studies have more directly tapped motor pathways in perception tasks. For example, performance in phoneme-monitoring paradigms (i.e., a speech perception task where participants identify whether a word contains a particular phoneme) is affected when participants maintain a motor plan to produce speech that contains the target phoneme (Roelofs, Özdemir, & Levelt, 2007). Another study showed that silent articulation (e.g., mouthing /ka/) while listening to speech (e.g., acoustic /pa/) can alter perception (e.g., perceive /ka/) (Sams, Möttönen, & Sihvonen, 2005). Finally, TMS studies have also shown that auditory perception of speech is affected by repeated stimulation of speech-motor cortex, and that this perceptual effect is selective to the specific part of motor cortex (i.e., the lip or tongue area) that is stimulated (D'Ausilio et al., 2009; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Möttönen & Watkins, 2009).

These data again suggest close links between motor and perceptual systems, and support the notion that information about the movements of articulators provide the basis for perceiving and categorizing speech information (Best, 1993; Browman & Goldstein, 1992; Fowler & Galantucci, 2005). However, several controversies still remain. It has been argued, for example, that tasks purporting to show the influence of motor information in perception are confounded with spreading activation to non-motor processes, which may instead influence later stages of perceptual processing (Lotto et al., 2009; Mahon & Caramazza, 2008). For example, perceptual effects reported in TMS studies that stimulate motor cortex might still be explained by auditory feedback from abstract conceptual or decision-making processes, which could be activated in parallel with motor stimulation (see discussion linked online to D'Ausilio et al., 2009). Thus, while there is consensus among researchers that articulation is implicated in auditory speech

perception, the precise nature of this relation remains unknown: are motor processes the source of modulation in speech perception, or do articulatory influences originate instead from parallel priming of associated auditory categories?

Furthermore, there is some controversy over the precise characteristics of motor processes that may contribute to perceptual analysis. Motor movements—articulatory movements in the domain of speech—are often accompanied by somatosensory feedback about the positions of individual effectors. Somatosensory information about articulator positions can be used, for example, to guide the movements of the articulators in achieving a particular speech target (e.g., say “seeb”) (Tremblay, Shiller, & Ostry, 2003). Information about whether or not the target has been achieved can be extracted from somatosensory information alone, independently of auditory feedback: this is shown clearly, for example, in the articulatory motions of adults with impaired hearing (Nasir & Ostry, 2008). Furthermore, it has also been argued that somatosensory feedback influences the perception of speech (Ito, Tiede, & Ostry, 2009; Nasir & Ostry, 2009). In a compelling example of this phenomenon, Ito et al. (2009) reported that perceptual categorization of /a/ and /e/ vowels were altered when an externally controlled mechanical arm deformed the skin and facial muscles of a listener in ways similar to what usually happens when these vowels are produced. Thus, these results suggest that “motor” influences in perception may actually be somatosensory in nature.

The field is currently debating the precise role that motor information (and/or the somatosensory inputs that results from motor movements, or what will be collectively be called “sensorimotor” information) plays in the perception of auditory speech. This debate is exemplified quite clearly in the studies reported by Sams et al. (2005), who showed that concurrent articulation influences the perception of speech. These results can be broadly interpreted in two ways: first, they may show that sensorimotor

information is the source of perceptual modulation. Alternatively, they may show that planning and executing an articulatory gesture will prime abstract auditory representations. Sams et al. (2005) interpreted their results by postulating that a forward model in the articulatory-motor system generates an “efferent copy” of the motor command, which is sent in parallel to speech effectors and to auditory cortex (p. 433). These efferent copies activate “phoneme-specific” information in auditory cortex in anticipation that perceptual systems will imminently receive auditory input that matches the information contained in the original motor command. Thus, the source of perceptual modulation, at least on the account advocated by Sams et al. (2005), is activation of an abstract phonemic category that is activated in parallel with a speech motor command.

In summary, the literature on articulatory influences in speech perception supports a divergent set of theoretical claims. On the one hand, some interpretations of the empirical data have suggested that articulatory influences are derived directly from sensorimotor information, either from the activation of articulatory-motor movements themselves (Fowler & Galantucci, 2005), or from somatosensory feedback related to the execution of these movements (Ito et al., 2009). On the other hand, others have suggested that planning and executing a motor command leads to the activation of abstract, phoneme-specific representations (Sams et al., 2005). In order to distinguish between these possibilities, we modified the behavioural paradigm used by Sams et al. (2005) to test whether articulatory effects on perception are attributed to executing movements (i.e., sensorimotor modulation) or to auditory and/or phonological representations activated in parallel with these movements (i.e., auditory imagery).

2.1 Experiment 1

Sams et al. (2005) reported that articulating one syllable while listening to another syllable can bias perceptual identification of the heard-syllable. Our own pilot

data replicated these results with a new set of disyllables. Silently mouthing /ava/ while listening to acoustic /aba/ induces particularly strong misperception of an illusory /ava/. Experiment 1 used this illusion to test a dissociation of the effects derived from sensorimotor information when articulating /ava/ (i.e., articulator movement) and associated auditory imagery that accompanies articulation (i.e., priming of phonemic categories).

Participants identified disyllables of naturally produced /aba/ and /ava/ tokens in two separate blocks with four conditions per block. In the Articulate Block, one condition was an auditory baseline where participants simply listened to and identified /aba/ and /ava/ speech stimuli. In three other conditions participants also identified /aba/ and /ava/ stimuli, but silently articulated one of three disyllables in synchrony with the auditory stimuli: either /aba/, /ava/, or /afa/. Articulation of /aba/ and /ava/ while listening to acoustic /aba/, for example, replicates the influence of matching versus mismatching articulation on the speech percept, originally tested by Sams et al. (2005). Articulation of /afa/, however, provides a critical test of whether misperception is modulated by sensorimotor aspects or auditory imagery. Making this speech gesture recruits the same articulators as when producing /ava/, but is acoustically and phonologically distinct when spoken aloud due to salient differences in voicing (i.e., vocal fold vibration). During silent articulation, however, voicing output differences are neutralized, creating a case where sensorimotor aspects of articulation remain similar, but corresponding auditory representations associated with the motor command differ. If sensorimotor aspects, rather than auditory priming, are what modulate misperception of acoustic /aba/ as illusory /ava/, then similarly high rates of /aba/ misperception are predicted in the /ava/ and /afa/ conditions compared to the /aba/ condition within the Articulate Block.

To further test whether the sensorimotor aspects of articulation modulate perception, rather than the internal auditory processes that accompany articulation, participants completed a second block: the Imagine Block. This block had an identical baseline condition as the Articulate Block (i.e., simply listening) and three additional conditions where participants *imagined* saying /aba/, /ava/, or /afa/ in synchrony with each auditory stimulus. Auditory imagery elicits patterns of cortical activation very similar to that elicited by actual perception (Kosslyn, Ganis, & Thompson, 2001). This task was thus designed to test whether auditory priming from imagining /ava/, /afa/, or /aba/ would result in misperception. If it is the sensorimotor information from executing movements that modulates misperception in the Articulate Block, then a different pattern of results is predicted in the Imagine Block: high rates of /aba/-misperception in the /ava/ condition and lower rates of misperception in the /afa/ and /aba/ conditions. Table 1 illustrates the predictions for each experimental condition within each block.

Table 2.1 – Predicted patterns of (mis)perception for acoustic /aba/ in Experiment 1.

The acoustic stimulus should be perceived as either /aba/ or an illusory /ava/ in the Articulate and Imagine blocks.

Condition	/aba/	/afa/	/ava/
Articulate Block	aba	ava	ava
Imagine Block	aba	aba	ava

2.1.1 Method

2.1.1.1 Participants

Forty-eight English-speaking undergraduates (28 female, $M = 22.8$ years, $SD = 6.6$ years) participated in this experiment. Total experiment time was about 25 minutes for each participant.

2.1.1.2 Stimuli

Five tokens each of /aba/ and /ava/ were recorded by a male speaker. Durations of the first vowel and consonant were adjusted using a waveform editor to equate for natural durational differences in the articulation of these sounds. The total duration of all tokens averaged 630 ms ($SD = 1.3$ ms).

This editing was done for two reasons that were related to the experimental procedure. First, participants were asked to rhythmically produce articulations along with the disyllables and consistent durational differences would require that subjects also adjust the rate of their articulations differently for acoustic /aba/ versus /ava/ tokens. Second, durational cues about a particular stimulus in our experiment were readily available to participants before perceptual judgments could be collected (see the footnote in the Procedure section below). Equating /aba/ and /ava/ durations ensured that participants would judge the identity of the spoken disyllables using the spectral characteristics of the acoustic signal, rather than these durational cues.

2.1.1.3 Procedure

Participants sat in a sound-attenuated booth in front of a computer monitor and keyboard, which were both connected to a Mac desktop computer placed outside of the booth and that was running experimental software (i.e., PsyScope). They were instructed to listen to presented speech sounds, and identify them as either /aba/ or

/ava/ by pressing one of two buttons on the keyboard. During one block, they were asked to silently articulate something in synchrony with presented speech sounds (i.e., the Articulate Block) and during another they were asked to imagine saying something in synchrony with the speech sounds (i.e., the Imagine Block).

Two aspects of an experimental trial facilitated synchronization of articulation or imagined articulation with auditory stimuli. First, each speech token was paired with a synchronous visual display where the diameter of a red circle on a white background tracked the amplitude of the auditory waveform. This visually highlighted dynamic changes in the auditory signal, making it easier to pace one's own articulations. This visual display and its corresponding auditory stimulus were synchronized to create a "target movie." Second, the structure of experimental trials encouraged participants to rhythmically produce articulatory movements before presentation of the target movie, thus improving synchronization. To accomplish this, "murmur movies" were created: low-pass filtered (145 Hz) versions of each auditory token, which eliminated spectral cues to the disyllable identity while preserving prosodic information, were paired with the same visual display as the corresponding target movie². A trial began with a key-press followed by a 500 ms delay, then three repetitions of a murmur movie followed by presentation of the corresponding target movie at an interstimulus interval of 300 ms. In the experimental conditions, participants were instructed to articulate or imagine saying something in synchrony with both the murmur and target movies, establishing a rhythm that co-occurred with the changing visual display. Sound levels were between 56 – 60 dB and between 64 – 68 dB for the murmur and target movies, respectively.

² This is also why durations of /aba/ and /ava/ tokens were equated: durational cues would still be present in low-pass filtered murmurs, and hearing these murmurs could have cued participants to the correct answer before target presentation.

After the target movie was presented, a prompt appeared and participants identified the target movie as either /aba/ or /ava/. Each participant completed twenty randomly ordered trials (each auditory token presented twice) per condition. The side of the /aba/ and /ava/ responses (i.e., “1” or “3” on the keypad), the order of all conditions within a block, as well as the order of the blocks were counterbalanced across participants.

2.1.2 Results

Misperception Indices (MPIs) for each acoustic target (i.e., /aba/ and /ava/) and for each experimental condition (i.e., /aba/, /ava/, and /afa/) within each block (i.e., Articulate or Imagine) were calculated for each participant by subtracting the proportion of correct-identifications in each condition from the corresponding block’s baseline proportion. A positive MPI (maximum = +1.0) indicates more misperceptions compared to baseline, and a negative MPI (minimum = -1.0) indicates fewer misperceptions. Preliminary analysis showed that MPIs for acoustic /aba/ differed significantly as a function of condition, while those for acoustic /ava/ did not in either the Articulate Block, $F(2, 94) = 1.14$, $p = .33$, $\eta^2 = .010$, or the Imagine Block, $F(2, 94) = .51$, $p = .60$, $\eta^2 = .004$. This was likely due to the fact that identification of (edited) /ava/ tokens were more ambiguous overall, obscuring differences between articulatory conditions. For example, the proportion-correct scores in the baseline condition within each block were lower for /ava/ ($M_{\text{articulate}} = .78$, $SD = .22$; $M_{\text{imagine}} = .79$, $SD = .21$) than for /aba/ ($M_{\text{articulate}} = .84$, $SD = .16$; $M_{\text{imagine}} = .89$, $SD = .11$). The MPIs for acoustic /ava/ and their standard deviations are displayed in Appendix A.

Subsequent analyses focused only on MPIs for acoustic /aba/, and participants who scored less than 50% when identifying /aba/ in either of the baseline conditions were eliminated from analysis ($n = 3$, 1 female). The remaining 45 participants showed

similarly high proportions of correct /aba/-identification in the baseline conditions from both the Articulate ($M = .91$, $SD = .13$) and Imagine ($M = .92$, $SD = .11$) blocks. MPIs to acoustic /aba/ from each experimental condition were analyzed in a 2 x 3 repeated-measures ANOVA (Block [Articulate or Imagine] x Condition [/aba/, /afa/, /ava/]). A significant interaction was observed, $F(2, 88) = 4.11$, $p = .020$, $\eta_G^2 = .012$, indicating that the pattern of MPIs differed as a function of articulating /aba/, /ava/, and /afa/ sounds versus imagining them (see Figure 2.1).

A follow-up ANOVA within the Articulate Block indicated that MPIs were not equivalent across conditions, $F(2, 88) = 6.66$, $p = .002$, $\eta_G^2 = .069$. Pair-wise comparisons (Bonferroni-corrected) confirmed our first predicted pattern of results: Acoustic /aba/ was misperceived as the percept /ava/ when articulating /ava/ significantly more than when articulating /aba/, $t(44) = 3.15$, $p = .009$, $d = .46$. Crucially, articulating /afa/ had an effect similar to articulating /ava/, resulting in more misperceptions compared to articulating /aba/, $t(44) = 3.51$, $p = .003$, $d = .51$.

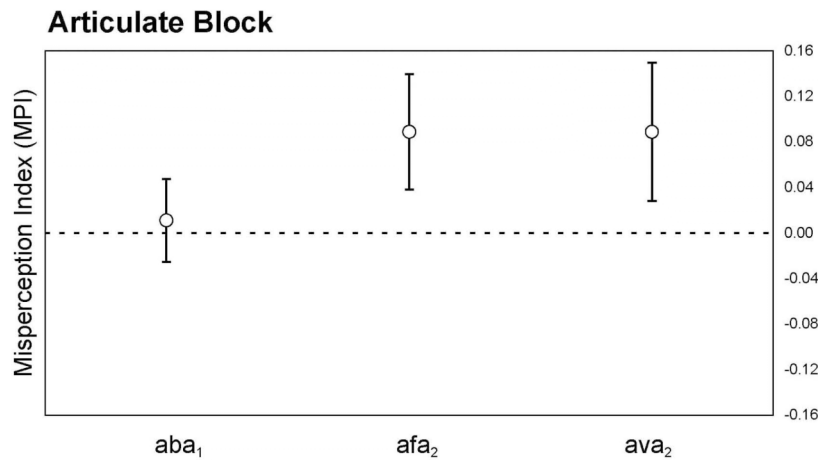
Another follow-up ANOVA on the Imagine Block also indicated that MPIs were not equivalent across conditions, $F(2, 88) = 6.95$, $p = .002$, $\eta_G^2 = .064$. Pair-wise comparisons confirmed our second prediction: Acoustic /aba/ was misperceived as the percept /ava/ when participants imagined saying /ava/ significantly more than when imagining /afa/, $t(44) = 2.70$, $p = .029$, $d = .39$, or when imagining /aba/, $t(44) = 3.26$, $p = .007$, $d = .47$.

Finally, Bonferroni-corrected 95%-confidence intervals of the MPIs within each block show that both /ava/ and /afa/ conditions resulted in significantly more misperceptions compared to the baseline condition in the Articulate Block, $t_{/ava/}(44) = 3.78$, $p = .001$, $d = .54$; $t_{/afa/}(44) = 4.60$, $p < .001$, $d = .65$. However, only the /ava/ condition showed this pattern in the Imagine Block, $t(44) = 3.15$, $p = .009$, $d = .46$.

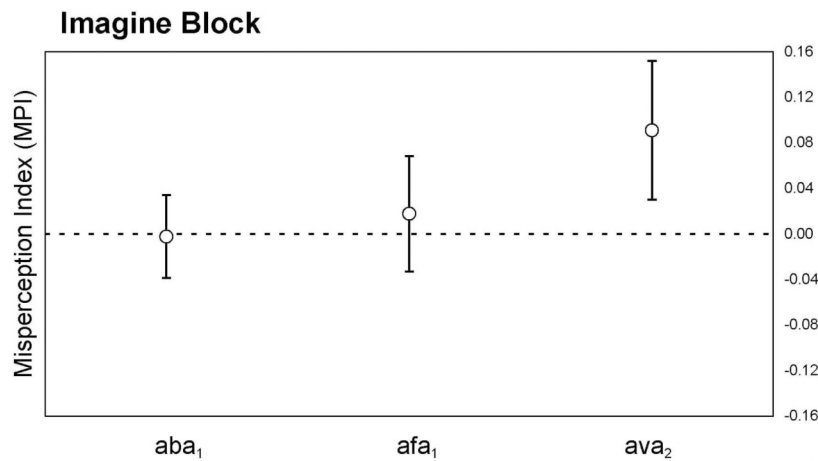
Figure 2.1 – Results for the Articulate and Imagine blocks in Experiment 1.

(A) Misperception indices (MPIs) for acoustic /aba/ in the ‘articulate’ block. (B) MPIs for acoustic /aba/ in the ‘imagine’ block. Error bars are Bonferroni-corrected 95% CIs. Common subscripts indicate conditions not significantly different from each other within the same block (i.e., pair-wise comparisons; Bonferroni-corrected alpha = .05).

A



B



2.1.3 Discussion

In Experiment 1, participants identified tokens of naturally produced /aba/ and /ava/ disyllables while silently and synchronously articulating, or while imagining

speaking aloud. Results showed that /aba/ perception was differentially affected in the various experimental conditions, although /ava/ perception was not. Subsequent analysis, which focused only on misperception of acoustic /aba/, revealed informative patterns. Results from the Articulate Block showed that silent articulation of /ava/-like movements (i.e., articulating either /ava/ or /afa/) while hearing acoustic /aba/ resulted in increased perception of an illusory /ava/ compared to a baseline condition. MPis when articulating /ava/ and /afa/ were also increased compared to when articulating of /aba/. Critically, these results show that articulating both /ava/ and /afa/ caused similar magnitudes of misperception, and suggest that the similar sensorimotor information (but different auditory associations) conveyed in the articulation of these disyllables contributed to misperception of acoustic /aba/.

Results from the Imagine Block further confirmed that auditory imagery associated only with /ava/ resulted in misperceptions above baseline. Moreover, the MPis while imagining /ava/ were greater in comparison to imagining either /aba/ or /afa/. Thus, auditory imagery associated with /ava/ and /afa/ had demonstrably different effects. Together, this shows that articulating both /ava/ and /afa/ while listening to acoustic /aba/ lead to similar magnitudes of misperception, and this happened despite the corresponding difference in the efficacy of auditory /ava/ and /afa/ imagery. These results show that sensorimotor aspects of articulation, rather than activation of auditory representations associated with it, can modulate speech perception.

2.2 Experiment 2

The Articulate Block from Experiment 1 showed that silent and synchronous articulation of /ava/ and /afa/ alter perception of acoustic /aba/ compared to baseline levels (i.e., just listening to acoustic /aba/), and also when compared to silent and synchronous articulation of /aba/. Experiment 2 was designed to add to these results in

two ways: controlling for a possible confound, and more directly testing strong versions of auditory versus sensorimotor accounts of our results above.

One possible confound in Experiment 1 is that the design allowed for the possibility that articulating *anything* while listening to acoustic /aba/ resulted in some instability when processing speech. Articulating /aba/ would have biased participants in the same direction as the acoustic stimulus, while the other articulatory conditions would have lead to a generalized decrease in perceptual accuracy of acoustic /aba/, no matter what participants were articulating. On this account, a similar facilitative effect for perception of acoustic /ava/ when articulating /ava/ is also predicted, but this effect may have been obscured by the low rates of correct /ava/ perception in the baseline condition. Thus, Experiment 2 was designed to test the alternative hypothesis: were our results in Experiment 1 caused by generalized perceptual instability related to articulation? To test this possibility, we replicated the critical effect found in the articulate /afa/ condition from Experiment 1, and also added distractor conditions that were dissimilar from either /aba/ or /ava/ targets (i.e., articulate /ama/ or /aya/). If the effect of perceptual instability were a legitimate confound, then articulating /afa/ should not result in a greater degree of misperception than articulating either /ama/ or /aya/ distractor conditions.

To more directly compare auditory versus sensorimotor accounts, a second experimental question was built into the design of Experiment 2. If actively articulating results in enhanced auditory priming compared to simply imagining articulation, then a purely auditory account may still be able explain the results from Experiment 1. For example, auditory /afa/ representations may indeed prime illusory /ava/ representations, but this priming relation might simply have been too weak to notice in the Imagine Block and strong enough to influence perception in the Articulate Block. To address this

possibility in Experiment 2, we added another condition where participants articulated /aðə/ (i.e., the initial consonant in “though”). Classic work on confusability matrices suggest that the consonant /v/ is acoustically confusable with the consonant /ð/, while relatively distinct from either /m/ or /f/ (e.g., Miller & Nicely, 1955). Thus, on a purely auditory account, articulating /aðə/ is predicted to result in significantly more misperceptions than either /afa/ or /ama/ (or, for that matter, /aya/), since articulating this consonant would activate auditory representations that are most easily confusable with the illusory /ava/ target. Articulating any of the other sequences would result in equivalent, but lesser degrees of perceptual influence. In other words, one would expect to find the following order of decreasing MPis on this account: /aðə/ > /afa/ = /ama/ = /aya/.

Predictions from a purely sensorimotor account of our results are straightforward: only /afa/ articulation would be predicted to have any perceptual influence, since it has an almost identical articulatory profile as the illusory target when silently articulating, while the other articulatory conditions (i.e., articulate /aðə/, /ama/, or /aya/) all differ in varying degrees. One would expect to find the following order of decreasing MPis on this account: /afa/ > /aðə/ = /ama/ = /aya/.

2.2.1 Method

Fifty English-speaking undergraduates (34 female; $M = 20.7$ years, $SD = 2.3$ years) were tested in a single Articulate Block using the same stimuli and procedure from Experiment 1. This block contained a baseline condition plus four additional conditions where participants silently articulated each of the following: /afa/ (i.e., a replication of the critical condition from Experiment 1), /aðə/, /ama/, and /aya/.

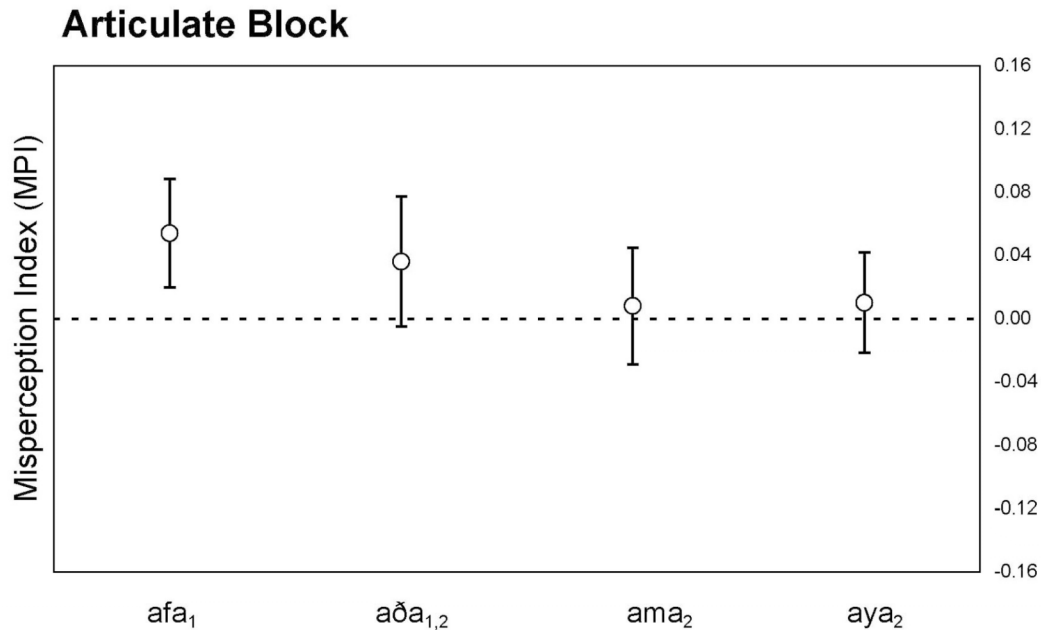
2.2.2 Results

As in Experiment 1, MPIs for perception of acoustic /ava/ showed no effects of condition, $F(3, 147) = .74, p = .53, \eta^2_G = .008$. Again, this may have been due to the inherent ambiguity of /ava/ tokens, as the mean proportion of correct identifications in the baseline condition was slightly lower for /ava/ tokens ($M = .89, SD = .16$) than for /aba/ tokens ($M = .95, SD = .10$). As in Experiment 1, only MPIs for acoustic /aba/ were subsequently analyzed. The MPIs for acoustic /ava/ and their standard deviations are displayed, along with those from Experiment 1, in Appendix A.

All participants met the 50% exclusion criterion used in Experiment 1. A repeated-measures ANOVA indicated that MPIs were not equivalent across conditions, $F(3, 147) = 4.22, p = .007, \eta^2_G = .037$. Pair-wise comparisons (Bonferroni-corrected) confirmed that articulating /afa/ resulted in significantly more misperceptions than articulating either /ama/, $t(49) = 3.05, p = .023, d = .42$, or /aya/, $t(49) = 3.09, p = .020, d = .43$. Misperception was intermediate for /aǎa/, which did not statistically differ from other conditions (see Figure 2.2). In addition, Bonferroni-corrected t-tests of the MPIs within each condition show that articulating /afa/ significantly influenced patterns of misperception compared to the baseline condition, $t(49) = 4.29, p < .001, d = .58$. Results also showed, however, that there was only a marginal difference in rates of misperception between the baseline condition and articulating /aǎa/, $t(49) = 2.31, p = .023, d = .32$.

Figure 2.2 – Results for the Articulate block in Experiment 2.

Error bars are Bonferroni-corrected 95% CIs. Common subscripts indicate conditions not significantly different from each other (i.e., pair-wise comparisons; Bonferroni-corrected alpha = .05).



2.2.3 Discussion

The design of Experiment 2 addressed two questions about our previous results. First, the pattern of /aba/ identifications shows that perceptual influences observed in Experiment 1 were *not* caused by a generalized instability in perception when articulating something besides the acoustic target. Results show that identification of acoustic /aba/ while articulating either /ama/ or /aya/ distractor conditions did not change from baseline, and also that *more* misperceptions resulted when articulating /afa/ than when articulating either the /ama/ or /aya/ distractors. This demonstrates that the perceptual influences observed in Experiment 1 were tied specifically to characteristics of the participants' articulatory patterns, and the similarity of those patterns to the articulation of the illusory /ava/ target.

Second, Experiment 2 was also designed to contrast strong versions of auditory versus sensorimotor accounts of our results. Recall that if *auditory* confusability were the sole source of perceptual influence, then one would have expected to find the following order of decreasing MPis: /aða/ > /afa/ = /ama/ = /aya/. This differs from the empirical results. Recall, again, that if a complete match of *sensorimotor* information were the sole source of perceptual influence (i.e., a purely sensorimotor account), then one would have expected to find the following order of decreasing MPis: /afa/ > /aða/ = /ama/ = /aya/. This provides a closer approximation of these results, but still does not explain the intermediate levels of misperception when articulating /aða/.

Another possibility, of course, is that these results show dual influences of sensorimotor and auditory information when articulating. That is, articulating /afa/ contributed to misperception of acoustic /aba/ due to its sensorimotor match to the illusory /ava/ target, while articulating /aða/ contributed to misperception (perhaps to a slightly smaller degree) due to its auditory similarity to the illusory target. Such a hybrid interpretation of our results would predict some degree of perceptual influence from articulation of /afa/ and /aða/, but little to no influence from articulation of either /ama/ or /aya/. This prediction maps well onto our empirical results.

In summary, Experiment 2 adds to our previous results in two specific ways. First, it rules out a possible confound in Experiment 1, which was related to the specificity of articulatory interference in perception. Second, results also argue against two theoretical accounts of our results: neither a purely auditory, nor a purely sensorimotor account generate predictions that match our empirical results. Results show that *both* sensorimotor and auditory aspects have detectable and separable influences on perception.

2.3 General Discussion

The nature of articulatory influences on auditory speech perception is the focus of current controversies in the field of speech research (Lotto et al., 2009; Massaro & Chen, 2008; Scott et al., 2009) and embodiment more generally (Mahon & Caramazza, 2008). Previous studies have shown that the motor system is somehow implicated in perception, but the precise characterizations of these action-perception linkages are considered controversial. One issue of particular importance concerns the *source* of articulatory influences: from what are the presently observed patterns of auditory (mis)perception derived? Previous work has suggested that these effects may be derived from an abstract auditory representation that is activated in parallel with a motor command (Sams et al., 2005), while other theories would suggest that sensorimotor information about the movements of the articulators can be a distinct source of perceptual modulation (D'Ausilio et al., 2009; Fowler & Galantucci, 2005).

This study was designed to provide evidence that might help in resolving this controversy. Results from Experiment 1 demonstrate that perceptual influences on speech perception can stem from information about the movements of individual articulators, rather than abstract auditory information (i.e., imagery) associated with a motor command. We find both types of influences in Experiment 2: misperception is derived, from articulating something that activates auditory information acoustically confusable with the illusory target (i.e., articulating /aõa/), and from making articulations that activate sensorimotor, but not auditory information, which is compatible with the illusory target (i.e., articulating /afa/). These results argue against the view that abstract phoneme categories, which are specified by a motor command, are the sole source of articulatory modulation in speech perception. Rather, our results suggest that

information from a variety of sources when articulating can provide a source of perceptual modulation, including sensorimotor aspects of articulation.

Importantly, the term “sensorimotor” has been used here to include the possibility that the perceptual modulation observed presently is derived from afferent (i.e., in-flowing) somatosensory feedback about the motions made by the articulators (Ito et al., 2009). This hypothesis is further supported by several recent studies that have offered striking demonstrations of speech-related somatosensory or tactile information influencing auditory speech perception (Fowler & Dekle, 1991; Gick & Derrick, 2009; Gick, Jóhannsdóttir, Gibrael, & Mühlbauer, 2008). Alternatively, perceptual modulation may derive from efferent (i.e., out-flowing) information about the coordinative dynamics of speech articulators (Saltzman & Kelso, 1987) or gestural scores (Browman & Goldstein, 1990, 1992), which are essentially abstractly defined relations between different speech articulators that determine how speech gestures should be executed. Further research, of course, must parse the precise meaning of sensorimotor influences, as well as the neural pathways that are involved in these distinct processes.

Two additional caveats merit brief mentioning: first, if it is the case that both auditory and sensorimotor information contribute to perceptual processing in speech, then why in Experiment 1 does articulating /ava/ merely result in a *similar* magnitude of misperception as articulating /afa/, which only shares sensorimotor information with the illusory percept? One possibility, of course, is that detecting more fine-grained differences between experimental conditions is not possible in this behavioural paradigm. Another possibility is that the mechanisms of sensorimotor and auditory modulation are simply not processes that are additive. This may indicate that the distinct pathways that underlie each type of perceptual influence are somehow incommensurable, or act independently in perceptual processing.

Second, our results support both strong and weak interpretations of motor theories of speech perception. A strong motor theory position would suggest that our results are not surprising, since perceptual identification simply involves identifying articulatory-motor information in the first place: that is, perceiving speech *is about perceiving articulation* (Fowler, 1986; Liberman & Mattingly, 1985; Liberman & Whalen, 2000). A weaker interpretation, one that we favour, suggests instead that sensorimotor information is just *one* source of perceptual influence, and may feed into a distributed network that *is* a percept. On this view, information across modalities, including sensorimotor, visual, and auditory information may be mutually excitable in such a way that suggests high degrees of interconnectivity (Pulvermüller, 2005).

In summary, our results are aimed at advancing the study of embodiment by investigating the source of articulatory influences on the perception of speech. Results suggest that articulation may lead to activation of both auditory and sensorimotor information, and both processes can be a source of perceptual modulation. Ultimately, results suggest that multiple sources of information about articulation, including sensorimotor aspects, can modulate speech perception. Future work must further establish what kinds of processes sensorimotor and perceptual systems share, and how connections between these systems are instantiated in the brain.

2.4 References

- Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. F. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life*, NATO Science Series (Vol. 69, pp. 289-304). Norwell, MA: Kluwer.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange & J. J. Jenkins (Eds.), *Speech perception and linguistic experience: Issues in cross-language research*, Cross-language speech perception (pp. 171-204). Timonium, MD: York Press.
- Browman, C. P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18(3), 299-320.
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, 49(3-4), 155-80.
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19(5), 381-385. doi:10.1016/j.cub.2009.01.017
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149-179.
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15(2), 399-402.
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14(1), 3-28.
- 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), 816-823.
- Fowler, C. A., & Galantucci, B. (2005). The relation of speech perception and speech production. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 633-652). Hoboken, NJ: Wiley.
- Galantucci, B., Fowler, C. A., & Goldstein, L. (2009). Perceptuomotor compatibility effects in speech. *Attention, Perception & Psychophysics*, 71(5), 1138-1149. doi:10.3758/APP.71.5.1138
- Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, 462(7272), 502-504. doi:10.1038/nature08572
- Gick, B., Jóhannsdóttir, K. M., Gibrael, D., & Mühlbauer, J. (2008). Tactile enhancement of auditory and visual speech perception in untrained perceivers. *The Journal of the Acoustical Society of America*, 123(4), EL72-EL76. doi:10.1121/1.2884349

- Gordon, P. C., & Meyer, D. E. (1984). Perceptual-motor processing of phonetic features in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 10(2), 153-178.
- Haruno, M., Wolpert, D. M., & Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Computation*, 13(10), 2201–2220.
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213-1216. doi:10.1126/science.279.5354.1213
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1245-1248. doi:10.1073/pnas.0810063106
- Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *NeuroImage*, 14(1), S103–S109.
- Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, 2(9), 635-642. doi:10.1038/35090055
- Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America*, 98(23), 13464-13471.
- 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
- Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4(5), 187 - 196. doi:10.1016/S1364-6613(00)01471-6
- Lotto, A. J., Hickok, G., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13(3), 110-114. doi:10.1016/j.tics.2008.11.008
- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, 102(1-3), 59-70. doi:10.1016/j.jphysparis.2008.03.004
- Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisited. *Psychonomic Bulletin & Review*, 15(2), 453-457; discussion 458-462.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-8. doi:1012311
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17(19), 1692-1696. doi:10.1016/j.cub.2007.08.064
- Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, 29(31), 9819-9825. doi:10.1523/JNEUROSCI.6018-08.2009
- Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*, 11(10), 1217-1222. doi:10.1038/nn.2193

- Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), 20470-20475.
- Prinz, W., & Hommel, B. (2002). *Common mechanisms in perception and action*. New York: Oxford University Press.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576-582. doi:10.1038/nrn1706
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351-360. doi:10.1038/nrn2811
- Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., 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
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169-92. doi:15217330
- Roelofs, A., Özdemir, R., & Levelt, W. J. M. (2007). Influences of spoken word planning on speech recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 900-913. doi:10.1037/0278-7393.33.5.900
- Saltzman, E. L., & Kelso, J. A. (1987). Skilled actions: A task-dynamic approach. *Psychological Review*, 94(1), 84-106. doi:10.1037/0033-295X.94.1.84
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, 23(2-3), 429-435.
- Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action - candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, 10(4), 295-302. doi:10.1038/nrn2603
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2006). Lending a helping hand to hearing: another motor theory of speech perception. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 250 - 285). Cambridge, UK: Cambridge University Press.
- Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, 423(6942), 866-869.
- Watkins, K. E., Strafella, A. 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
- 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), 701-702. doi:10.1038/nn1263

Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 107(2), 592-597.
doi:10.1073/pnas.0904774107

3: MAINTAINING A SINGLE ARTICULATORY POSITION CAN INFLUENCE SPEECH PERCEPTION³

Some theories have suggested that the perceptual analysis of events or actions can be influenced by representations or processes linked to the motor system (e.g., Haruno, Wolpert, & Kawato, 2001; Jeannerod, 2001; Prinz & Hommel, 2002; Rizzolatti & Craighero, 2004). This idea is echoed in speech research, where motor theories of speech perception have long suggested that articulatory processes are similarly recruited in the perceptual analysis of speech (Fowler, 1986; Galantucci, Fowler, & Turvey, 2006; Liberman & Mattingly, 1985).

Evidence for this claim comes from several types of speech perception tasks. For example, planning a spoken utterance can affect how quickly one can identify whether a word-label contains a particular sound (i.e., phoneme-monitoring) (Roelofs, Özdemir, & Levelt, 2007). In another paradigm, learning to modify the way that one articulates a particular sound can also modify the perception of these sounds in a subsequent listening-only task (Nasir & Ostry, 2009). Recent work has also shown that silently and simultaneously articulating while perceiving speech (e.g., mouthing /ka/ while hearing acoustic /pa/) can result in perceptual assimilation of the articulatory target (e.g., one perceives /ka/) (Sams, Möttönen, & Sihvonen, 2005; see also Chapter 2). Finally, several trans-cranial magnetic stimulation (TMS) studies have also shown that auditory speech perception is affected by repeated TMS to speech-motor cortex, and that this perceptual effect is selective to the specific part of motor cortex (i.e., the lip or tongue

³ A version of this chapter will be submitted for publication. Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2010) Maintaining Static Articulator Positions Influences Speech Perception.

area) that is stimulated (D'Ausilio et al., 2009; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Möttönen & Watkins, 2009).

These data suggest close links between articulatory processes and speech perception, but much remains unclear about how articulation modulates or influences perception. Indeed, articulating speech involves a cascade of distinct sensory and motor processes, and current debate has offered varying interpretations of the pathways responsible for influencing perception (Lotto, Hickok, & Holt, 2009; Massaro & Chen, 2008; Pulvermüller & Fadiga, 2010; Scott, McGettigan, & Eisner, 2009). Recent evidence has suggested that sensorimotor information about speech-like articulatory motions can be a distinct source of perceptual modulation (i.e., Chapter 2), but even so, little research has investigated how qualitatively different levels of information in speech motor control are integrated with speech perception. The current study is designed to address this question.

Researchers have identified at least two levels in the hierarchy of speech motor control: “task-level” information specifies the goals of articulation (i.e., the production of particular phonemes) and abstracts over the detailed movements of the speech articulators. Articulator-level information, in contrast, coordinates the moment-to-moment particulars of the positions and trajectories for the lips, tongue, jaw, etc. (Gracco, 1994; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; Saltzman & Munhall, 1989; Shaiman & Gracco, 2002). This is shown most clearly in studies that deliver mechanical perturbations to the jaw while experimental participants speak. The motor system maintains the dynamically coordinated structure of articulator movements that enables production of the speech target (i.e., task-level information) by making swift adjustments to the positions, force, and velocity of individual articulators (i.e., articulator-level information). This hierarchical nature of speech motor control is further reinforced in

transfer-of-learning tasks that also use these perturbation techniques: learned adjustments to a simple articulator movement (i.e., jaw-lowering when saying “ee-aa”) are not transferred when the same movement is embedded in a different speech context (e.g. jaw-lowering when saying “uu-aa”) (Tremblay, Houle, & Ostry, 2008).

Linguistic theories relating speech perception with speech production refer to abstract constructs that are more reminiscent of task- than articulator-level information. Gestural scores, for example, capture rich information about the coordinated dynamic relations between speech articulators, and also play a central role in theories of articulatory phonology (Browman & Goldstein, 1990, 1992). This is similarly the case in motor theories of speech perception, where it has always been assumed that detailed information about the coordinative structures of speech gestures are the basis of auditory speech perception as well as the basis of the link between speech perception and production (Fowler, 1996; Fowler & Galantucci, 2005; Galantucci et al., 2006).

To date, little empirical evidence has supported the claim that task-level information, or even speech-specific movements for that matter, is required to demonstrate an articulatory influence on speech perception. Some relevant data, however, come from a study showing that perceptual categorization of /a/ and /e/ vowels was altered when an externally controlled mechanical arm deformed facial skin near the mouth in a simple downward or upward motion, which approximates the deformations of the skin and muscles when /a/ and /e/, respectively, are produced (Ito, Tiede, & Ostry, 2009). This afferent (i.e., in-flowing) somatosensory feedback about articulator motion had to be somewhat specific to speech: “twitches” that deformed skin in a non-speech manner (i.e., at a faster rate), or simply statically holding skin in an “upward” or “downward” position had non-significant effects on perception. While this suggests that highly impoverished information about articulator movements can still influence speech

perception, the conveyed articulatory information must retain at least some dynamic similarities to actual speech gestures.

There are several reasons, however, to question the conclusion that dynamic articulatory information is necessary to observe perceptual influences. First, visual inspection of Ito et al.'s published data suggest that holding skin in static positions may indeed have influenced perception, but that the effect was not significant due to lack of statistical power, or because Ito et al. (2009) looked only at identification scores rather than more sensitive measures like reaction time. Second, low-level articulatory information in efferent (i.e., out-flowing) channels may have a different effect on perception than afferent (i.e., in-flowing) ones. In other words, different kinds of information may be fed to the perceptual system when the motor system is actively maintaining a single articulatory position by sending information through efferent channels compared to when the perceptual system is simply receiving dynamic and speech-like afferent somatosensory feedback from skin and muscle receptors.

In order to evaluate these possibilities, experimental participants in the current study were instructed to maintain a single articulatory position that was embedded in a non-speech gesture (i.e., holding one's breath) while making speeded identifications of speech syllables. If it were shown that the maintenance of this articulatory position could still influence speech perception, this would suggest that task-level, dynamic, and speech-specific information about articulation is not the only source of perceptual influence from processes associated with speech production. Rather, the locus of perceptual influence may stem from more basic information about the positions of the articulators within the vocal tract.

3.1 Method

Articulations of the stop consonants /p/, /t/, and /k/ mutually involve obstructing and then releasing airflow through the vocal tract (i.e., they are all voiceless stops). These consonants crucially differ, however, in the mechanism by which airflow is stopped (i.e., their places of articulation): /p/ is produced by closing the lips, /t/ by placing the tongue tip on the alveolar ridge, and /k/ by raising the tongue body against the velum (i.e., the soft palate). The current experiment asked whether maintaining an articulatory position that obstructed airflow at a specific place of articulation (i.e., holding one's breath at the lips) could modulate perception of related consonants (i.e., /p/ versus /t/ or /k/).

In one experimental block participants were presented with an acoustically edited continuum between /pa/ and /ta/, and were asked to classify each token. Both identification and reaction times were recorded, and this was repeated in a second block, where participants classified another continuum between /ta/ and /ka/. During each of these two blocks, participants held their breath in short spurts, sometimes at their lips (as if “diving underwater”), which is related to /p/ articulation, and other times at their glottis (as if “lifting a heavy object”), which is not related to either /p/, /t/, or /k/ articulation. This design allowed us to ask whether maintaining a static articulatory position (i.e., holding one's breath) could influence the perception of speech sounds that are related to that articulatory position. In other words, it was predicted that performance in perceptual tasks related to /pa/ identification would be selectively affected by holding one's breath at the lips.

Previous research has suggested that articulatory influences generally result in perceptual facilitation or assimilation, such that auditory percepts are biased towards the speech percepts that are congruent with articulatory information (D'Ausilio et al., 2009;

Ito et al., 2009; Möttönen & Watkins, 2009; see also Chapter 2). Thus, we predicted more /pa/-biased perceptions of the /pa/-/ta/ continuum when breath holding at the lips versus at the glottis. Alternatively, we predicted no differences between the lips and glottis conditions when perceiving /ta/-/ka/. Furthermore, reaction times (i.e., decision latencies) may also be influenced differently in lips versus glottis conditions. The direction of this influence is not as clearly predicted, however: reaction times when classifying a /pa/-/ta/ continuum may be facilitated only for /pa/ and delayed for /ta/, or alternatively they may be facilitated across the board (i.e., deciding between the /pa/ or /ta/ may be easier when a /pa/ information is pre-potently activated). In either case, some difference between reaction times between /pa/-/ta/ versus /ta/-/ka/ continua is predicted in the lip condition, but not in the glottis condition. A table summarizing these predictions is provided below.

Table 3.1 – Predicted results listed by the corresponding breath holding condition.

Breath Holding	Lips	Glottis
Identification of /pa/-/ta/	/pa/-bias	no bias
Identification of /ta/-/ka/	no bias	no bias
Decision Latencies	/pa/-/ta/ \neq /ta/-/ka/	/pa/-/ta/ = /ta/-/ka/

3.1.1 Participants

Eight native English-speakers ($M = 26.4$ years, $SD = 7.5$ years) were recruited from a pool of volunteers who were willing to participate in behavioural experiments for money (i.e., \$10 per hour), or for course credit in a psychology course.

3.1.2 Stimuli

Three naturally produced tokens of /pa/, /ta/, and /ka/ from a male speaker were edited to equate for duration, and entered into an electronic software package (STRAIGHT; see Kawahara, Masuda-Katsuse, & de Cheveigné, 1999). This package produced two separate six-member continua of speech sounds (i.e., /pa/-/ta/ and /ta/-/ka/ continua). The six tokens on the /pa/-/ta/ continuum contained 0.29%, 35.92%, 46.84%, 51.44%, 68.1%, and 93.97% of the information supplied by the /ta/ end-point, while the remainder of this information came from the /pa/ end-point (see Kawahara et al., 1999 for details). The six tokens of the /ta/-/ka/ continuum similarly contained 14.08%, 50%, 62.93%, 67.53%, 81.9%, and 99.71% of the information supplied by the /ka/ end-point. All of these values were determined based on extensive pre-testing with a separate group of native English speakers. The duration of the tokens used in the /pa/-/ta/ continuum were within 2 ms of each other, averaging 341 ms. The duration of the tokens used in the /ta/-/ka/ continuum were also similar in range, averaging 340 ms.

3.1.3 Procedure

Participants were instructed that they were participating in an experiment that explored the effect of physiological stress on auditory perception. This ensured that participants would be less likely to draw the connection between breath holding at the lips and /p/ production. Indeed, none of the participants made this connection when debriefed at the end of the experiment. During the instruction period, breath holding at the lips and glottis was explained and demonstrated by the experimenter, who also ensured that the mouth was left open when holding one's breath at the glottis. Moreover, the experimenter ensured that participants felt the difference at the designated location in the vocal tract when holding their breath at the lips versus the glottis (i.e., participants were asked if they could feel pressure at the base on their throat when holding breath at

the glottis, but not at the lips). Participants also demonstrated breath holding in front of the experimenter, who ensured that they were following instructions as carefully as possible.

Testing took place in a sound-attenuated booth, and stimuli were presented over computer speakers at approximately 67 dB using PsyScope software loaded onto a Mac computer. Participants sat in front of a monitor and responded on a button box, completing experimental trials that were structured to allow participants to engage in periodic breath holding. For example, a prompt on the screen informed subjects that they should begin holding their breath at the designated place of articulation. Once their breath was held, participants were instructed to press the spacebar, and an experimental stimulus was presented after 500 ms. Participants pressed one of two keys to identify this stimulus as accurately and as quickly as possible: a visual prompt remained onscreen to remind participants which key to push. Once a response was made, another stimulus was presented after a delay of 500 ms, and another response was recorded. This happened a total of 3 times before a screen appeared, which indicated that it was time to release one's breath. When ready, participants could press the spacebar and three trials were again presented. Participants were given the opportunity to practice doing this before beginning the actual experiment until they felt comfortable with the task.

Participants completed 36 trials with one type of breath holding (i.e., lips or glottis) before an onscreen prompt indicated that the style of breath holding should be switched. Lips and glottis trials for one continuum (e.g., /pa/-/ta/) alternated in this manner for 4 cycles (i.e., 288 trials total). Each token of the six-member continuum was presented a total of 24 times, equally in the first, second, or third slot in the triads of trials during which breath was held. Once finished, the experimenter entered the booth to

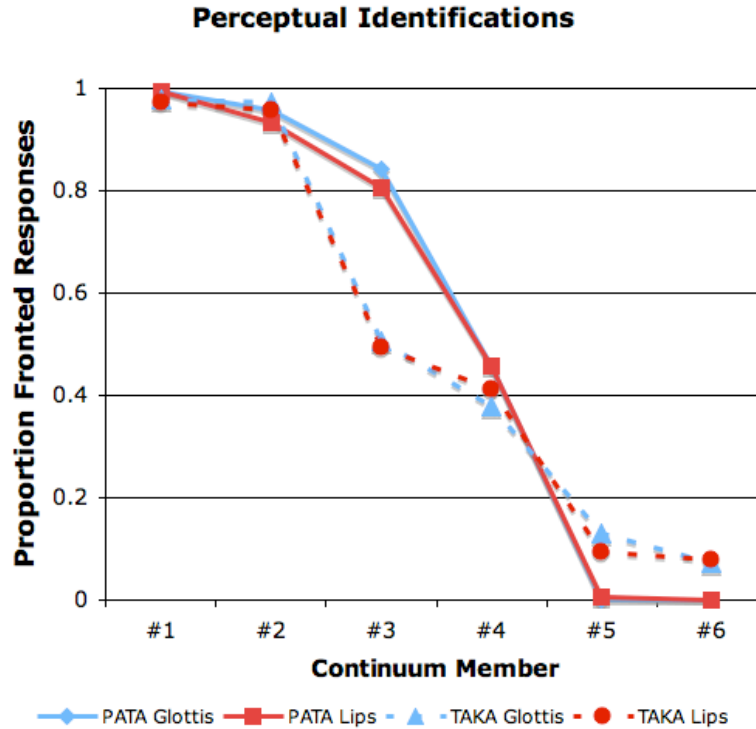
explain that the next block of the experiment would involve perceptual identification of another continuum (e.g., /ta/-/ka/), which was presented in a similar manner. The side of the responses (i.e., “left” or “right” keys on the button box), whether the first breath holding condition was at the lips or the glottis, and the order of the two continuum blocks (i.e., /pa/-/ta/ or /ta/-/ka/) were fully counterbalanced across individuals in the study. The whole experiment took about 45 minutes to complete.

3.2 Results

Identification responses for both the /pa/-/ta/ and /ta/-/ka/ continua were recoded as the proportion of “fronted” responses, which refers to the place of articulation closer to the front of the mouth (i.e. /pa/ for the /pa/-/ta/ continuum, and /ta/ for the /ta/-/ka/ continuum). The mean proportions of fronted responses are illustrated in Figure 3.1, which suggests that breath holding at the lips or the glottis did not differentially affect speech categorization. This was confirmed by a repeated measures ANOVA on the proportion of fronted responses, with factors of CONTINUUM (i.e., /pa/-/ta/ or /ta/-/ka/), PLACE of breath holding (i.e., lips or glottis), and MEMBER of the continuum (i.e., #1-#6, from furthest front to furthest back in the mouth). As expected, no significant interactions involving the PLACE approached significance. There was, however, a marginal interaction between the factors of CONTINUUM and MEMBER, $F(5, 35) = 2.43$, $p = .054$, $\eta_G^2 = .13$, indicating that the boundary between /pa/ and /ta/ categories was in a slightly different place on that continuum than the corresponding boundary on the /ta/-/ka/ continuum (see Figure 3.1). In addition, there was a significant main effect of MEMBER, indicating, as expected, that the proportion of fronted responses differed within both continua, $F(5, 35) = 143.69$, $p < .001$, $\eta_G^2 = .81$. In summary, these results did not support the prediction that maintaining a static articulatory position (i.e., breath holding at the lips) could have an influence of perceptual identifications of /pa/ syllables.

Figure 3.1 – Results for perceptual identifications.

Responses for the /pa/-/ta/ (solid) and /ta/-/ka/ (dashed) continua are plotted for each continuum member, where #1 indicates the place of articulation furthest to the front of the mouth (i.e., /pa/ for /pa/-/ta/ and /ta/ for /ta/-/ka/) and #6 indicates the place of articulation furthest to the back (i.e., /ta/ or /ka/). Colors indicate results for the lip (red) and glottis (blue) breath holding conditions.



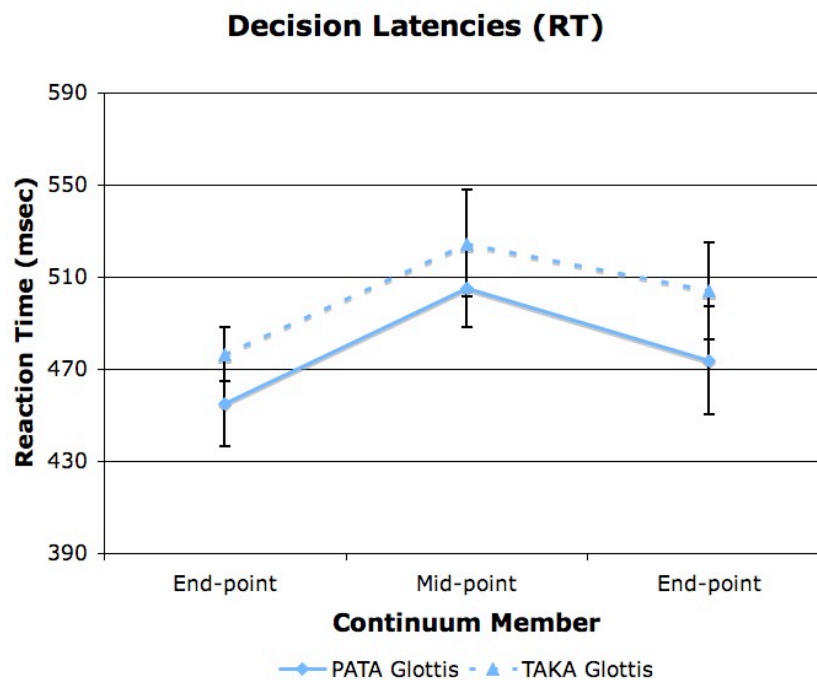
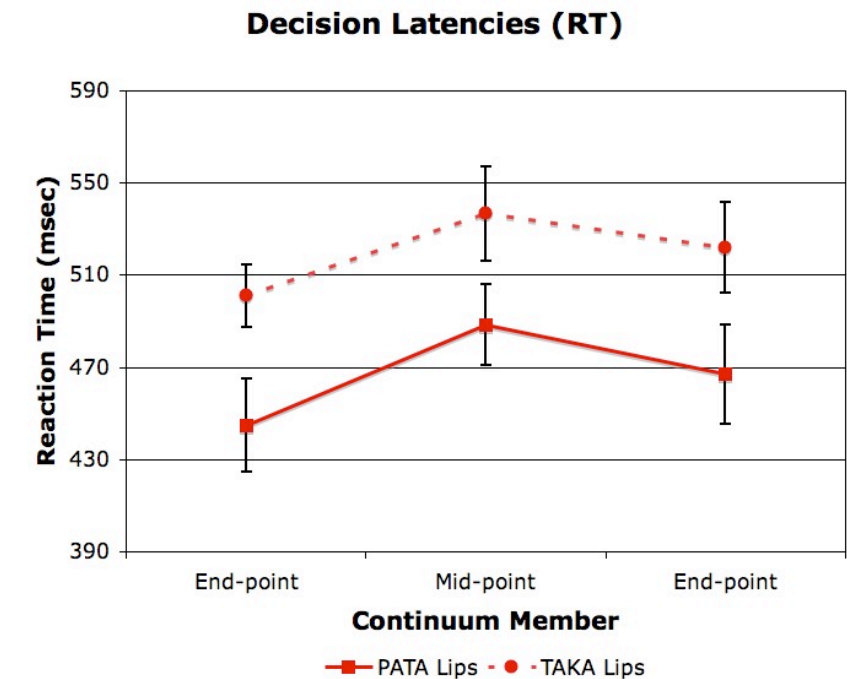
While perceptual identifications were not differentially affected by the breath holding conditions, subsequent analysis on the reaction times suggested that breath holding did influence decision latencies. Preliminary processing eliminated reaction times that were shorter than 150 ms or longer than 1500 ms, which were *a priori* thresholds designed to eliminate responses that were too short or too long to reflect actual decision processes. Less than .01% of the data were eliminated in this step. A repeated measures ANOVA was conducted on the reaction times with factors of CONTINUUM (i.e., /pa/-/ta/ or /ta/-/ka/), PLACE (i.e., lips or glottis), and MEMBER (i.e., #1-#6). Results again showed a main effect of MEMBER, $F(5, 35) = 14.46$, $p < .001$, η_G^2

= .14, which reflects the fact that participants in categorical perception studies usually take longer to categorize mid-points on a continuum compared to the end-points, due to the fact that the middle tokens are more perceptually ambiguous.

Interestingly, this ANOVA also revealed a significant main effect of CONTINUUM, $F(1, 7) = 6.76$, $p = 0.035$, $\eta_G^2 = .11$, and an interaction between PLACE and CONTINUUM, $F(1, 7) = 5.65$, $p = .049$, $\eta_G^2 = .018$, but no additional main effects or interactions. This suggests that reaction times in categorizing the /pa/-/ta/ continuum were shorter overall than those in the /ta/-/ka/ continuum, but that this difference interacted with the effects of holding one's breath at either the lip or the glottis. An analysis of the simple main effects derived from this interaction revealed that reaction times in the lip condition were faster when categorizing the /pa/-/ta/ continuum ($M = 467$ ms, $SD = 47$ ms) than when categorizing the /ta/-/ka/ continuum ($M = 520$ ms, $SD = 47$ ms), $t(7) = 3.70$, $p = .008$, $d = 1.31$. However, this was not the case within the glottis condition ($M_{/pa/-/ta/} = 578$ ms, $SD = 51$ ms; $M_{/ta/-/ka/} = 501$ ms, $SD = 49$ ms), $t(7) = 1.82$, $p = .22$, $d = .48$. These simple main effects, as well as the main effect of MEMBER, which illustrates differences between the mid- and end-points on the continuum, are displayed graphically in Figure 3.2.

Figure 3.2 – Results for decision latencies.

The /pa/-/ta/ (solid) and /ta/-/ka/ continua (dashed) are displayed for the lip (top in red) and glottis (bottom in blue) conditions. For simplicity in visual presentation, continuum members have been binned into two end-point bins (i.e., #1 and #2 *versus* #5 and #6), as well as a mid-point bin (i.e., #3 and #4). Error bars indicate standard error.



3.3 Discussion

Previous work has demonstrated that sensorimotor information can modulate speech perception, but little research has investigated *what* sorts of information in the motor system drives this effect. The current study addresses this question, asking whether articulatory influences necessarily originate from speech-specific information about the dynamically specified movements of the tongue, lips, jaw, etc. Results suggest that this is not the case: maintaining a single articulatory position (i.e., simply holding one's breath at the lips) can enhance the speed at which participants categorized tokens of related speech (i.e., /pa/ and /ta/ sounds) compared to decision latencies when classifying unrelated speech (i.e., /ta/ and /ka/ sounds). Specifically, our effects stemmed from the fact that the production of /p/ sounds is related to breath holding at the lips, as similar differences in decision latencies were not found when one held their breath at the glottis. This enhancement of decision latencies for the /pa-/ta/ continuum may reflect the fact that articulatory information when breath holding at the lips is activating /pa/-related processes in perceptual analysis, helping to speed decisions (i.e., either /pa/ or not /pa/) when identifying members of this particular continuum, but not influencing decision processes for the /ta-/ka/ continuum. Thus, results suggest that static maintenance of one's articulators in a relevant position, even when that position is embedded in a non-speech gesture, is enough to bias performance in a speech perception task.

These data challenge a previous claim, which suggested that speech-like deformations of facial skin and muscles can affect perception of vowels, but that static deformations have no statistically significant effects on perceptual categorization (Ito et al., 2009). Two major differences between that paradigm and our current data may help to explain these divergent results. First, our results showed an articulatory influence only

in the analysis of the reaction times to make categorical judgments, rather than in the analysis of the categorical judgments themselves. Only the latter was recorded previously. Future work is needed to establish how information about dynamic movements *versus* static positions may differ in their contribution to perceptual modulation. Second, the current study provided information about both efferent (i.e., out-flowing) motor information, as well as afferent (i.e., in-flowing) somatosensory information when maintaining lip closure during breath holding, while Ito et al.'s (2009) study was designed to provide only the latter. Future work will be needed to parse the potentially separate contributions of these distinct pathways on perception, examining how afferent somatosensory feedback differs when the articulators remain static as opposed to when they are moving.

Even though our design does not investigate the nature of speech production directly (rather, it investigates speech perception), the original motivation of the present study is crucially rooted in work on speech motor control, and our results address an important controversy in that field. Speech motor control research has suggested that motor control is organized hierarchically, and that information at the highest levels of this hierarchy specify “goals” or “tasks” that the articulators are meant to achieve. It has always been assumed that speech perception, insofar as it interacts with production, taps these higher levels. Importantly, however, the characterization of these high-level tasks in the speech motor control literature is controversial: for example, articulatory processes may be targeted at producing acoustic events (e.g., Houde & Jordan, 1998, 2002; Jones & Munhall, 2005; Perkell et al., 2000); they may instead be defined abstractly by dynamic coordinative relations between different speech articulators, (Browman & Goldstein, 1990, 1992; Kelso et al., 1984; Saltzman & Munhall, 1989); alternatively, speech motor control may be based on somatosensory feedback from

peripheral receptors in skin and muscle about the location and movement dynamics of the articulators (Nasir & Ostry, 2008; Tremblay, Shiller, & Ostry, 2003).

Two basic conclusions can be drawn from our results, vis-à-vis this controversy in speech research. On the one hand, our results suggest one way in which motor control might be structured. Specifically, because neither auditory information nor dynamic coordinative relations were involved in the production side of our task, our data tentatively suggest that speech motor control is at least partially defined by the achievement of simple articulatory positions. This remains speculative, however, since the present data come from a perception task, rather than a motor control task. On the other hand, our work more firmly establishes that low-level representations or processes within the speech motor control hierarchy (i.e., information about static articulatory positions, even in non-speech gestures) are accessible at the interface between motor control and perceptual processing. Strikingly, this is counter to the general assumption in speech research that low-level information in speech motor control does not play an important role in perceptual processing.

In summary, our results suggest that several levels in the hierarchy of speech motor control may interact with speech processing, including low-level information about the positions of the articulators. This supports a view of speech perception where a wide variety of information across modalities, including sensorimotor, visual, and auditory information may be recruited in perceptual analysis (Pulvermüller, 2005). Specifically, our results suggest that the dominant hierarchical view of speech motor control is not strictly reflected in the link between speech perception and production. Rather, low-level articulatory information present when one simply holds their breath (i.e., a decidedly non-speech gesture) can selectively influence the perception of related consonants.

3.4 References

- Browman, C. P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18(3), 299-320.
- Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, 49(3-4), 155-80.
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19(5), 381-385. doi:10.1016/j.cub.2009.01.017
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14(1), 3-28.
- Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, 99(3), 1730-41. doi:8819862
- Fowler, C. A., & Galantucci, B. (2005). The relation of speech perception and speech production. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 633-652). Hoboken, NJ: Wiley.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361-377.
- Gracco, V. L. (1994). Some organizational characteristics of speech movement control. *Journal of Speech & Hearing Research*, 37(1), 4-27.
- Haruno, M., Wolpert, D. M., & Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Computation*, 13(10), 2201-2220.
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213-1216. doi:10.1126/science.279.5354.1213
- Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, 45(2), 295-310. doi:10.1044/1092-4388(2002/023)
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1245-1248. doi:10.1073/pnas.0810063106
- Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *NeuroImage*, 14(1), S103-S109.
- Jones, J. A., & Munhall, K. G. (2005). Remapping auditory-motor representations in voice production. *Current Biology*, 15(19), 1768-1772. doi:10.1016/j.cub.2005.08.063
- Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*, 27(3), 187-208.

- Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10(6), 812-832. doi:10.1037/0096-1523.10.6.812
- 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
- Lotto, A. J., Hickok, G., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13(3), 110-114. doi:10.1016/j.tics.2008.11.008
- Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisited. *Psychonomic Bulletin & Review*, 15(2), 453-457; discussion 458-462.
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17(19), 1692-1696. doi:10.1016/j.cub.2007.08.064
- Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, 29(31), 9819-9825. doi:10.1523/JNEUROSCI.6018-08.2009
- Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*, 11(10), 1217-1222. doi:10.1038/nn.2193
- Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), 20470-20475.
- Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., et al. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, 28(3), 233-272. doi:10.1006/jpho.2000.0116
- Prinz, W., & Hommel, B. (2002). *Common mechanisms in perception and action*. New York: Oxford University Press.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576-582. doi:10.1038/nrn1706
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351-360. doi:10.1038/nrn2811
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169-92. doi:15217330
- Roelofs, A., Özdemir, R., & Levelt, W. J. M. (2007). Influences of spoken word planning on speech recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 900-913. doi:10.1037/0278-7393.33.5.900
- Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1(4), 333-382.

- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, 23(2-3), 429-435.
- Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action - candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, 10(4), 295-302. doi:10.1038/nrn2603
- Shaiman, S., & Gracco, V. L. (2002). Task-specific sensorimotor interactions in speech production. *Experimental Brain Research*, 146(4), 411-418. doi:10.1007/s00221-002-1195-5
- Tremblay, S., Houle, G., & Ostry, D. J. (2008). Specificity of speech motor learning. *Journal of Neuroscience*, 28(10), 2426.
- Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, 423(6942), 866-869.

4: ACHIEVING LIP-SHAPES WHILE SUCKING AND CHEWING INFLUENCES INFANTS' AUDIOVISUAL SPEECH PERCEPTION⁴

4.1 Introduction

Theoretical approaches have long asserted that the perception and production of speech are closely linked (e.g., Fowler, 1986; Guenther, Hampson, & Johnson, 1998; Hickok & Poeppel, 2007; Liberman & Mattingly, 1985; Pulvermüller & Fadiga, 2010; Rizzolatti & Arbib, 1998; Skipper, Nusbaum, & Small, 2006). Emerging evidence has supported this assertion by demonstrating striking influences of speech perception-on-production and of speech production-on-perception. For example, hearing speech automatically and implicitly alters articulatory movements during speech production, even when the heard speech is irrelevant to the production task (Yuen, M. H. Davis, Brysbaert, & Rastle, 2010). Similarly, perceptual identification of speech is influenced by articulating (Nasir & Ostry, 2009; Sams, Möttönen, & Sihvonen, 2005), by transcranial magnetic stimulation of speech motor cortex (D'Ausilio et al., 2009; Möttönen & Watkins, 2009), and even by deforming facial skin and muscles in a way that simulates articulation (Ito, Tiede, & Ostry, 2009).

The ontogenetic origins of these links between speech perception and production are relatively unexplored. On the one hand, several reports suggest that speech perception can influence infants' earliest vocalizations: phonological patterns in one's native (or maternal) language can affect cry melodies in newborns (Mampe, Friederici,

⁴ A version of this chapter will be submitted for publication. Yeung, H. H. & Werker, J. F. (2010) Achieving Lip-shapes while Sucking and Chewing Influences Infants' Audiovisual Speech Perception.

Christophe, & Wermke, 2009), early vowel-like vocalizations (Ruzza, Rocca, Boero, & Lenti, 2006), vocalic and consonantal characteristics of babbling (de Boysson-Bardies, Halle, Sagart, & Durand, 1989; de Boysson-Bardies, Sagart, & Durand, 1984; Levitt & Utman, 1992; Whalen, Levitt, & Goldstein, 2007), as well as the earliest productions of words (de Boysson-Bardies & Vihman, 1991; McCune & Vihman, 2001; Vihman, 1991, 1993). Yet, on the other hand, no studies to date have systematically examined the effect of production-on-perception in infants (i.e., whether making articulatory movements can influence infants' perception of speech).

This gap in the developmental literature is related to asymmetries in the development of speech perception versus production: infants perceive sophisticated, language-specific phonetic patterns in speech long before a correspondingly sophisticated system of speech production is in place (see Locke, 1983; Oller, 1980; Stark, 1980; Werker & Tees, 1999, 2005 for reviews). This fact about development has made it difficult to postulate how speech production processes in infancy could influence speech perception (see, for example, Nazzi, Bertoncini, & Bijeljac-Babic, 2009). Indeed, only a few discussions have hypothesized links between articulatory processes and infant speech perception (Best, 1995; Kent & Vorperian, 2007; Kuhl & Meltzoff, 1984; Werker, 1993). For example, direct-realist accounts of perceptual development have suggested that amodal information about gestural events in the world provides the basis for infant speech processing (Best, 1995). A related theoretical approach has further postulated mappings between infants' own articulatory representations and speech information in both auditory and visual modalities (Kent & Vorperian, 2007; Kuhl & Meltzoff, 1984, 1988). This "articulatory mapping" hypothesis, especially, suggests a basis from which one might study the effects of speech production-on-perception in infancy.

The articulatory mapping hypothesis originated from work on cross-modal matching in speech: young infants hearing a vowel look more at a face visually articulating the matching vowel in a side-by-side visual display of two talking faces (Kuhl & Meltzoff, 1982, 1984; Kuhl, Williams, & Meltzoff, 1991; MacKain, Studdert-Kennedy, Spieker, & Stern, 1983; Patterson & Werker, 1999, 2003). Several lines of evidence suggest, at least indirectly, that articulatory mappings provide the basis for this matching. First, matching behaviour is specific to speech: infants do not match when spectral information in the vowels is removed (Kuhl & Meltzoff, 1984), or when tone complexes with similar acoustic characteristics as the vowels are played instead (Kuhl et al., 1991). Second, infants under 6 months of age detect cross-modal matches of non-native speech and faces, suggesting that these effects do not stem from purely audiovisual experience seeing adults produce these speech sounds (Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009; Walton & Bower, 1993). Third, infants often produce congruent mouth-shapes when participating in a cross-modal matching paradigm, and this does not happen when hearing non-speech tones instead of vowels (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999). Finally, a few studies have observed vocal imitation in more experimental contexts, suggesting that infants are able to link talking faces with their own vocal configurations (Chen, Striano, & Rakoczy, 2004; Kessen, Levine, & Wendrich, 1979; Kuhl & Meltzoff, 1996; Legerstee, 1990).

There are two major limitations to the hypothesis that infant cross-modal speech perception has an articulatory basis. First, previous work has never provided direct evidence for this by manipulating infants' articulatory behaviour and evaluating the effect of this behaviour on perception. The need for such evidence has been highlighted by recent reports, which show that temporal synchrony, rather than articulatory information, forms the basis of many cross-modal phenomena in the infant speech perception

literature (Lewkowicz, 2000, 2010). While the temporal synchrony between the speech and faces is well controlled in most cross-modal matching studies (Kuhl & Meltzoff, 1982, 1984; MacKain et al., 1983; Patterson & Werker, 1999, 2003), it is possible that infants use other types of dynamic information inherent in even the most well-matched audiovisual displays. Specific evidence is needed that that an articulatory process, *per se*, is involved.

A second limitation to the articulatory mapping hypothesis is that it remains unclear *how* articulatory information should be represented in young infants. The literature on developmental speech motor control has suggested several alternatives with respect to whether early orofacial movements are linked with actual speech production at older ages. Some research has suggested, for example, that babbles may be an important unit of analysis, since babbling is shaped by speech perception and argued to be continuous with first productions of words (de Boysson-Bardies & Vihman, 1991; McCune & Vihman, 2001; Vihman, 1991). Other research argues instead that many aspects of babbling reflect universal constraints on the motor system that may not be specific to speech (B. L. Davis & MacNeilage, 1995; MacNeilage & B. L. Davis, 1993). A third possibility, which is based on analysis of the coordinative movements made by different groups of jaw muscles when infants speak, babble, suck, or chew, suggests that motor mechanisms underlying these other orofacial movements are not continuous at all with later speech motor control (Moore & Ruark, 1996; Steeve, Moore, Green, Reilly, & McMurtrey, 2008).

Several experiments are presented here, which address these two limitations to the articulatory mapping hypothesis, and directly test the broader question of whether infants' articulatory-motor movements can influence speech perception. Experiments 1 and 2 establish a baseline, replicating and extending previous work on cross-modal

matching that has presented [i]- and [u]-faces⁵ to 4.5-month-olds, an age commonly tested in this paradigm (Baier, Idsardi, & Lidz, 2007; work cited in Kuhl & Meltzoff, 1988). In Experiment 3, we manipulate the types of articulatory movements made by infants tested in the same experimental paradigm. Specifically, we utilized the similarity in lip-shapes achieved when adults articulate these particular vowels and lip-shapes achieved when infants engage in certain non-speech behaviours. For example, lip-spreading when articulating /i/ is similar to lip-shapes achieved when infants mouth certain objects (i.e., ones wider than the mouth); lip-rounding when articulating /u/ is similar to lip-shapes achieved when infants suck on a pacifier or fingertip.

If results in Experiments 1 and 2 differ from those in Experiment 3, this would show that executing a particular type of articulatory movement could influence the perception of speech in a cross-modal matching task, addressing the first issue above. This would also show that information from non-speech movements that mimic articulatory characteristics of speech (i.e., mouthing or sucking) could be integrated with speech information at some early point in development, thus shedding light on the second issue highlighted above. Together, these experiments would provide definitive evidence for an effect of articulatory information on infant speech perception, helping to characterize the ontogenesis of the link between perceptual and motor processes in this domain.

4.2 Experiment 1

Experiment 1 establishes infants' baseline preferences for [i]- and [u]-faces, which are used for the visual display in all subsequent experiments. A video containing a side-by-side display of a woman articulating [u] and the same woman articulating [i] (i.e.,

⁵ Brackets (i.e., [i]-faces) will be used to denote visual speech information, while slashes (i.e., /i/-sounds) will be used to denote auditory speech information.

“ee”) were shown to infants for a 2-minute period. While this video was shown, a synchronized audio track of the vowel /a/ (i.e., “aa”) was played, which matched neither of the displayed faces. We predicted that infants would show no systematic preference for either face.

4.2.1 Method

4.2.1.1 Participants

Sixteen infants (9 female) with an average age of 4 months 20 days ($R = 4;7 - 4;28$) were recruited from a database of families willing to participate in research. Infants heard mostly English in the home, as measured by parental report ($R = 65\% - 100\%$; $M = 95\%$).

4.2.1.2 Stimuli

Videos from Baier et al. (2007) were used, which consisted of side-by-side displays of [i]- and [u]-faces. The [i]-face appeared on the left side in one video, and on the right side in another. Each display was constructed from 20 clips (10 each) of [i]- and [u]-articulations, the onsets of which were synchronized and occurred every 2 seconds. In addition, the duration of mouth opening and the onset of blinking were both synchronized. Ten clip-sequences were looped until each video played continuously for approximately 2 minutes.

Stimuli videos were also presented with one of several audio tracks, which were recorded in a separate session by the same woman filmed for the video. During the speech recording session, the woman watched videos of herself articulating [a], [i], and [u] (i.e., the same clip-sequences as above) and produced the matching vowel as closely as possible with the original audio track. Ten tokens of the vowels /a/, /i/, or /u/ were selected from this recording session, and used to create a new audio track where the

onsets of the separately recorded vowels coincided with the vowel onsets in the original audio track. Durations of all the vowels were similar ($M_{/a/} = .44$ s; $M_{/i/} = .44$ s; $M_{/u/} = .52$ s). Durations of mouth opening in the [i]- and [u]-video displays were slightly longer than the vowels ($M_{/i/} = 1.36$ s; $M_{/u/} = 1.32$ s), which also corresponded closely to the actual temporal dynamics between face and voice in the original recordings.

4.2.1.3 Procedure

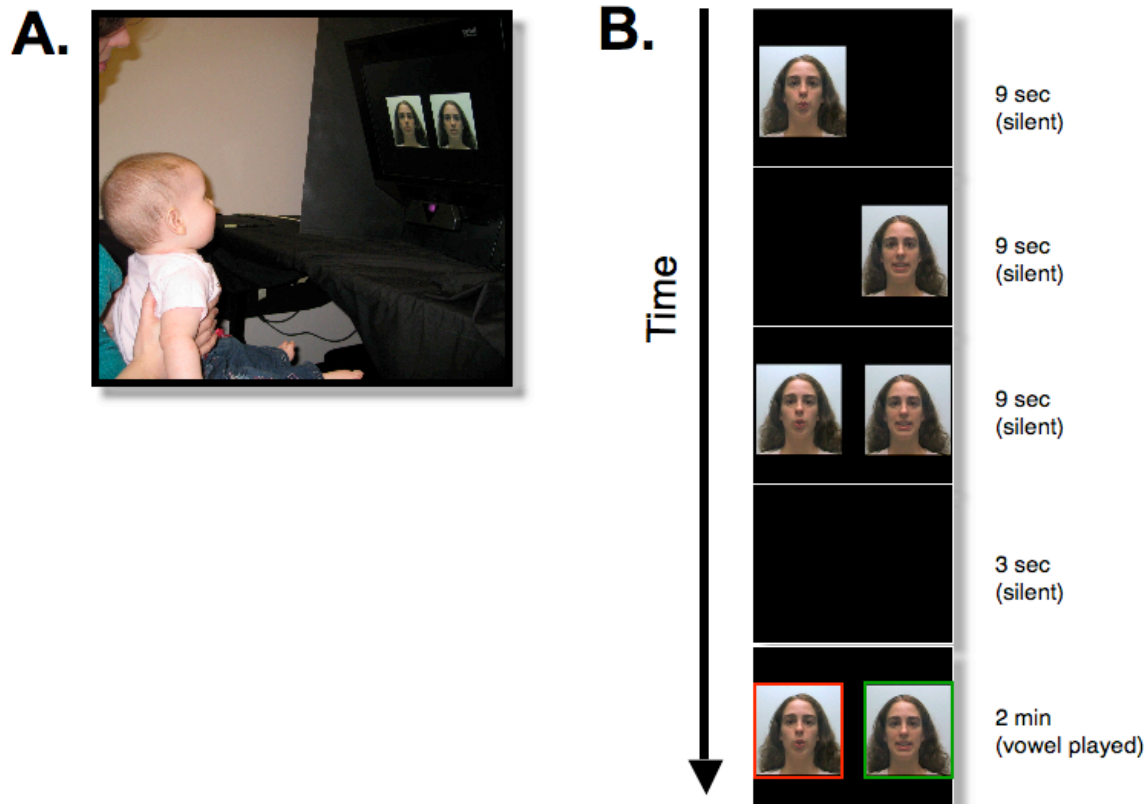
Infants were seated in their caregiver's lap while eye-gaze was recorded at a sampling rate of 50 Hz with a Tobii 1750 eye-tracker. The 38.1 cm eye-tracking monitor (i.e., 15" diagonal) was positioned approximately 60 cm from the head of the infant at a visual angle of approximately 30 degrees. As displayed on a black background, each of the two face-displays covered a 9.8 cm x 9.8 cm square, symmetrically oriented around the central point on the screen and separated horizontally by 2.7 cm. A small video camera was concealed approximately 30 cm to the bottom right of the eye-tracking monitor, and recorded the infant's face during the procedure. During the test video, sound pressure levels ranged between 60 - 64 dB, and sound emanated from two speakers bilaterally positioned behind a black cardboard barrier surrounding the eye-tracker.

Infants' gaze was calibrated immediately before the test session. A looming blue ball appeared sequentially in five places: the centre of the screen and each of its corners. Each time the looming ball appeared, it was accompanied by several beeping sounds to maintain attention. The infant's face was viewed through the video camera feed and the position of the calibrated point was marked when infants looked to be fixating on a corner or the centre location. This procedure took 1 - 2 minutes to complete. The test procedure began after the calibration, and closely followed previous paradigms (Patterson & Werker, 1999, 2003). One of the faces was displayed in silence

for 9 s and was followed by a 9 s display of the other face on the other side. Both faces were displayed in silence for another 9 s, and finally the screen went blank for 3 s before the 2 min test movie was played (see Figure 4.1). This familiarization procedure informed infants that a face would be appearing on each side of the screen before presentation of the test movie. Which side appeared first, as well as whether the /i/-face appeared on the left or right were counterbalanced. Caregivers were instructed to prevent chewing or mouthing on the hands and to focus on the infant in their laps rather than on the screen.

Figure 4.1 – An illustration of the apparatus and procedure.

(A) The experimental apparatus. (B) A schematic of the video timeline for one version of the procedure. The red and green squares indicate regions of interest used in the analysis of infants' gaze.



4.2.1.4 Analysis

Gaze analysis was conducted in two regions of interest that were of identical dimensions, and oriented over the two faces in the video display (see Figure 4.1). Raw gaze data were calculated over the 2 m that infants watched the test display; no fixation filters or interpolative calculations were applied, since we were interested in overall interest in each of the faces, rather than specific patterns of fixations.

In addition to the sample of sixteen infants, four other infants were tested, but excluded based on three *a priori* criteria derived from preliminary gaze analysis. First, infants were excluded if at least 4 calibration points were not recorded before the test session ($N = 0$). Second, the duration of recorded gaze at the faces must have exceeded at least 40 s in the 2-minute test video (i.e., 1/3 of the total time) ($N = 4$). Infants failed to meet this criterion if they were excessively fussy and disinterested in the video, or if they shifted their angle with respect to the screen so that the eye-tracker was unsuccessful at calculating gaze. Finally, infants could also be excluded if they looked less than 1 s at one of the two faces, demonstrating a side-bias ($N = 0$). This latter criterion was applied based on the assumption that these infants may have had trouble disengaging from one of the faces, and followed similar inclusion criterion used previously (Kuhl & Meltzoff, 1984; Patterson & Werker, 1999, 2003).

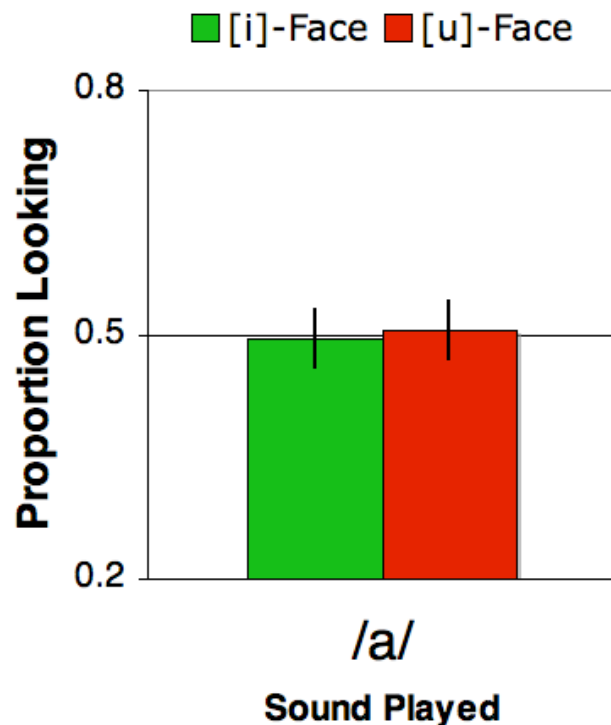
4.2.2 Results

Gaze data were recorded for an average of 62.2% ($SD = 14.2\%$) of the test video, and recalculated as a proportion-score to the [i]-face for each infant. Preliminary analysis also examined other factors that might have influenced looking patterns: both gender (i.e., male or female) and the side that the [i]-face appeared (i.e., left or right) were entered as factors into a between-subjects ANOVA on the proportion scores, but no interactions or main effects reached significance (i.e., $\alpha = .05$).

The main analysis showed that infants had no significant preference for either face, looking at the [i]-face for an average of 49.5% ($SD = 14.9\%$) of the time (see Figure 4.2), $t(15) = .13$, $p = .90$, $d = .033$. In addition, 6 of 16 infants looked longer at the [i]-face, which was not different from chance by a binomial test, $p = .23$.

Figure 4.2 – Proportion-looking at the [i]- and [u]-faces while hearing the vowel /a/ (Experiment 1).

The error bars indicate standard error.



4.2.3 Discussion

Experiment 1 showed that infants do not have a baseline preference for either the [i]- or [u]-face. Previous work showing infants a side-by-side display of these particular [i]- and [u]-faces has also reported a neutral preference during a silent familiarization period, which was similar in format to the first 3 trials used here (Baier et al., 2007). However, visual preferences in silence may not reflect visual behaviour when

auditory stimuli are presented, as suggested by work on auditory overshadowing in young infants (Sloutsky & Robinson, 2008). Thus, the present results confirmed that infants do not have a baseline preference for [i]- or [u]-faces even when an unrelated vowel (i.e., /a/) is presented over a synchronized audio track. This provides a foundation from which to interpret subsequent experiments.

4.3 Experiment 2

Given that infants have no baseline preferences for either [i]- or [u]-faces when listening to an unrelated vowel sound, Experiment 2 sought to replicate cross-modal matching behaviour when infants heard related /i/ and /u/ vowels (Baier, Idsardi, & Lidz, 2007; also, work cited in Kuhl & Meltzoff, 1988). The same video of a woman articulating [i] and [u] was shown. Unlike Experiment 1, however, a synchronized audio track of the vowel /i/ or the vowel /u/ was played, which matched only one of the displayed faces. Based on previous work, we predicted that infants would look longer to the matching face.

4.3.1 Method

4.3.1.1 Participants

Thirty-two infants (16 female) with an average age of 4 months 20 days ($R = 4;0 - 5;3$) were recruited as before. Infants heard mostly English, as measured by parental report ($R = 75\% - 100\%$; $M = 96\%$), and were randomly assigned to either the /i/- ($N = 16$; 8 female) or /u/-vowel ($N = 16$; 8 female) conditions. Eight additional infants were excluded due to procedural errors, including experimenter error ($N = 5$) and equipment failure ($N = 3$).

4.3.1.2 Stimuli

Stimuli were identical to those used in Experiment 1, except the audio tracks of the test displays contained either the vowel /i/ or /u/. The recording parameters of these stimuli have been described above.

4.3.1.3 Procedure

The procedure was identical to the one used in Experiment 1, except two new groups of infants were tested: one group heard the /i/ audio track, and another heard the /u/ audio track.

4.3.1.4 Analysis

The analysis was identical to the one used in Experiment 1, and in this experiment fifteen additional infants were excluded based on the three *a priori* criteria taken from preliminary analysis of the eye-tracking data: if less than 4 calibration points were recorded ($N = 1$); if recorded gaze summed to less than 40 s during the test video ($N = 13$); or if infants demonstrated a side-bias ($N = 1$).

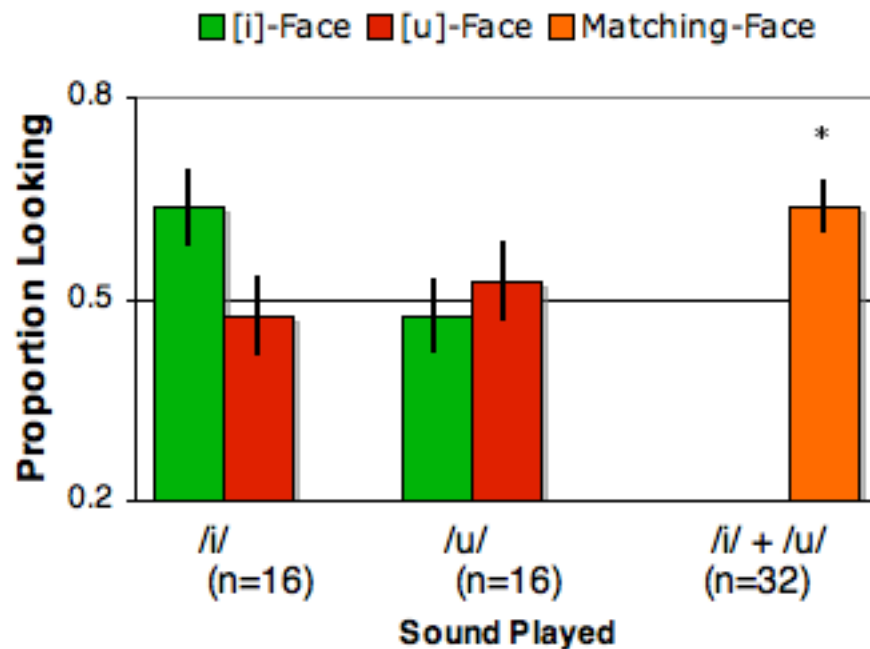
4.3.2 Results

Gaze data were recorded for an average of 64.3% ($SD = 19.4\%$) of the test video, and recalculated as a proportion-score for the matching face. Preliminary analysis examined several factors that might have influenced looking patterns: gender (i.e., male or female), the side that the matching-face appeared (i.e., left or right), and the matched-vowel (i.e., [i] or [u]) were entered as factors into a between-subject ANOVA on the proportion scores to the matching face. However, no interactions or main effects reached significance.

This was followed by the main analysis, which examined cross-modal matching in these infants. For infants who heard /i/, proportion-looking to the [i]-face averaged 63.6% ($SD = 22.8\%$); for the infants who heard /u/, proportion-looking to the [u]-face averaged 52.6% ($SD = 23.4\%$) (see Figure 4.3). Altogether, infants looked more at the matching face than chance (i.e., 50%), $t(31) = 1.97$, *one-tailed* $p = .029$, $d = .35$, replicating previous cross-modal matching studies using this paradigm (Baier et al., 2007; Kuhl & Meltzoff, 1982, 1984; Kuhl et al., 1991; Patterson & Werker, 1999, 2003). Additionally, 22 of 32 infants looked longer to the matching face, which was significantly more than predicted by chance by a binomial test, $p = .025$.

Figure 4.3 – Proportion-looking at [i]- and [u]-faces while hearing either /i/ or /u/ (Experiment 2).

The proportion of matching across both groups is shown in orange. The error bars are located on the matching proportion and indicate standard error. The asterisk indicates the significant effect of matching ($p < .05$, *one-tailed*).



4.3.3 Discussion

Infants this age have previously been shown to match /i/- and /u/-vowel sounds to [i]- and [u]-faces (Baier, Idsardi, & Lidz, 2007; also cited in Kuhl & Meltzoff, 1988). Experiment 2 replicates this effect, providing the foundation for subsequent work, which manipulates the articulatory movements that infants make while watching these faces.

4.4 Experiment 3

In the current experiment, we ask whether making articulatory-motor movements, which are relevant to articulations made in the audio and visual targets, will disrupt or affect the cross-modal matching behaviour observed in Experiment 2. Two specific types of motor movements were selected: lip-spreading and lip-rounding. These lip-shapes, of course, correspond with the visual [i] and [u] articulations as well as the auditory /i/ and /u/ vowels used in Experiment 2. The same paradigm as in Experiment 2 was presented to two groups of infants. The lip-spreading group was given a toy or finger on which to mouth during the study: these objects were oriented or sized in such a way that lips would need to be spread in order to accommodate it. The lip-rounding group was given a pacifier or fingertip on which to suck during the study: this similarly ensured that infants' lips would be rounded when sucking.

If achieving these lip-shapes affects infants' ability to match speech information across the auditory and visual modalities, then it is predicted that infants will show a different pattern of preferences than in Experiment 2. Three more specific predictions can be made. One possibility is that these infants will show an assimilation effect, such that the speech percept is biased towards or captured by articulatory-motor information, which echoes previous work in the adult literature (D'Ausilio et al., 2009; Ito et al., 2009; Möttönen & Watkins, 2009; Nasir & Ostry, 2009; Sams et al., 2005). It would be predicted that the lip-spreading group would be activating motor features shared with

audio /i/ and visual [i], facilitating the matching of /i/ vowels to [i]-faces on the one hand, and perhaps impairing the matching of /u/ vowels to [u]-faces on the other. This would result in an overall preference for the [i]-face, and no matching. Similarly, it would be predicted that the lip-rounding group would show an overall preference for the [u]-face, and again no matching.

Second and third possibilities are that achieving lip-shapes will result in an interference effect or a contrast effect. This is predicted by theories of action-perception from outside of the domain of speech, where it is thought that engaging motor processes sometimes withholds motor-related information from perceptual analysis, instead of facilitating the activation of shared representations (Hamilton, Wolpert, & Frith, 2004; Hommel, Bertoncini, Aschersleben, & Prinz, 2001; Müsseler & Hommel, 1997; Schütz-Bosbach & Prinz, 2007). Specifically, a general interference effect might be predicted if achieving lip-shapes withheld the use of articulatory mappings in cross-modal matching. In this event, infants would show only their baseline preferences observed in Experiment 1 (i.e. they would look 50% at either face). A contrast effect might be predicted if achieving lip-shapes selectively withheld articulatory information that is specific to one vowel-type: the lip-spreading group would be suppressing articulatory features shared with audio /i/ and visual [i] speech, which might facilitate matching /u/ vowels to [u]-faces. This would result in an overall preference for the [u]-face, and no matching. Similarly, it would be predicted that the lip-rounding group would suppress matching of /u/ vowels to [u]-faces, but facilitate matching /i/ vowels to [i]-faces, resulting in an overall preference for the [u]-face, and again no matching.

4.4.1 Method

4.4.1.1 Participants

Sixty-four infants (32 female) were recruited as before. The lip-spreading and lip-rounding groups had a similar average age ($R = 4;0 - 5;3$; $M_{spread} = 4;17$; $M_{round} = 4;16$). Infants heard predominantly English in the home, as measured by parental report ($R = 30\% - 100\%$; $M_{spread} = 83\%$; $M_{round} = 89\%$). Equal numbers of infants from each lip-shape category (balanced for gender) were randomly assigned to either the /i/-vowel ($N = 32$; 16 female) or /u/-vowel ($N = 32$; 16 female) conditions. Two additional infants were excluded due to experimenter error ($N_{spread} = 2$).

4.4.1.2 Stimuli

Stimuli were identical to those used in Experiment 2.

4.4.1.3 Procedure

The procedure was identical to the one used in Experiment 2, except that two new groups of infants were tested. Infants in the lip-spreading group ($N = 32$) chewed or mouthed part of a larger object (i.e., too large to be a choking-hazard), and infants' lips were typically spread to accommodate the object's width. Most infants chewed or mouthed a clean wooden teething ring⁶ provided by the experimenters (1.2 cm in thickness and 6.8 cm in diameter; see Fig. 4.4) ($N = 23$), but a handful of infants preferred another type of commercially available teething toy ($N = 7$, four of whom used "Sophie the Giraffe"⁷), the side of their parent's finger ($N = 2$), or a combination of these objects ($N = 2$). Lip-rounding was encouraged in another group ($N = 32$) by allowing infants to suck on part of an object. Usually this object was a pacifier ($N = 28$; see Fig.

⁶ Item #1004 manufactured by Camden Rose ®

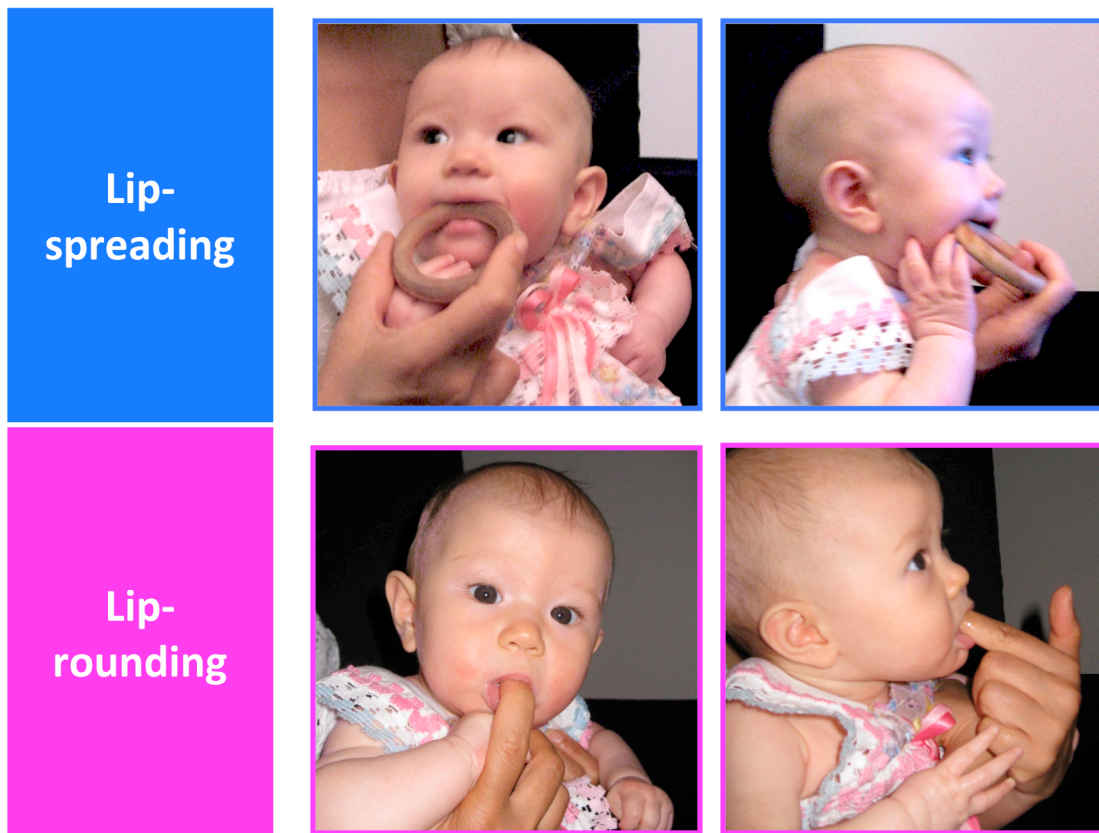
⁷ Manufactured by Vulli ®

4.4). For a few infants, however, either the tip of their caregiver's finger ($N = 3$), or a combination of the finger and the pacifier ($N = 1$) were placed in the mouth instead.

Caregivers were instructed to watch their infant during the procedure and prevent the finger, toy, or pacifier from being spit out, and adjust it if it was dislodged. In the event that an object fell out of their infant's mouth, caregivers were asked to replace it. Clean teething rings or pacifiers were available under the caregiver's chair for this purpose.

Figure 4.4 – Still images of an infant demonstrating lip-spreading and lip-rounding.

The teething ring pictured in the lip-spreading images is of the same model given to most infants in that group. Infants in the lip-rounding group mostly sucked on a pacifier, rather than a fingertip, but these images allow a view of the lips that would otherwise be concealed by the pacifier shield.



4.4.1.4 Analysis

The analysis was identical to the one used in Experiment 2. Nineteen additional infants were run, but excluded based on the three *a priori* criteria derived from preliminary analysis of the eye-tracking data: if less than 4 calibration points were recorded ($N_{round} = 1$); if recorded gaze summed to less than 40 s during the test display ($N_{spread} = 7$; $N_{round} = 8$); or if infants demonstrated a side-bias ($N_{spread} = 1$; $N_{round} = 2$).

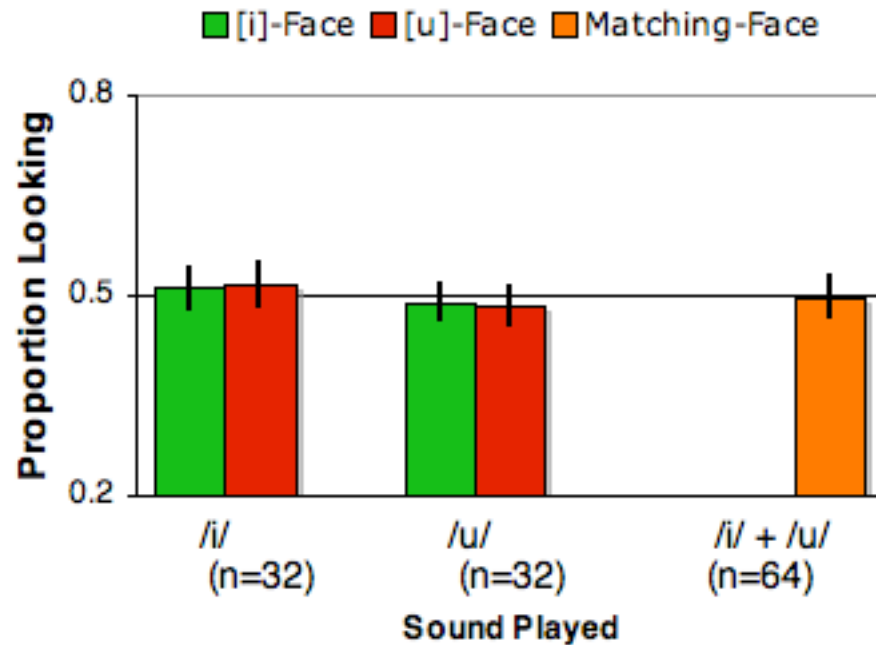
4.4.2 Results

Gaze data were recorded for an average of 69.1% ($SD = 18.1\%$) of the total test time. Infants in the lip-spreading group looked marginally less ($M_{spread} = 65.0\%$; $SD = 18.6\%$) than the lip-rounding group ($M_{round} = 73.1\%$; $SD = 16.9\%$), $t(62) = 1.84$, $p = 0.071$, $d = .46$. This was likely due to the fact that infants who were given a finger or pacifier on which to suck were somewhat less fussy or held at a different angle than infants who were given a toy or finger on which to chew.

As before, the gaze data to each of the faces were recalculated as a proportion-matching score. Overall, infants who heard /i/ looked at the matching face an average of 48.9% ($SD = 28.0\%$) of the time, while infants who heard /u/ looked time at the matching face an average of 51.5% ($SD = 24.6\%$) of the time (see Figure 4.5). Together, these results showed that infants did not look more than chance at the matching face, $t(63) = .071$, $p = .94$, $d = .009$, failing to replicate Experiment 2 as well as previous work that has shown cross-modal matching using this paradigm (Baier et al., 2007; Kuhl & Meltzoff, 1982, 1984; Kuhl et al., 1991; Patterson & Werker, 1999, 2003). Binary analysis of individual infants supported this result: only 34 of 64 infants looked longer to the matching face, which was not significantly more than predicted by chance by a binomial test, $p = .35$.

Figure 4.5 – Proportion-looking at [i]- and [u]-faces while hearing either /i/ or /u/ (Experiment 3).

The proportion of matching across both groups is shown in orange. The error bars are located on the matching proportion and indicate standard error.



To examine the possibility of assimilation or contrast effects (i.e., an interaction between lip-shape groups and matching performance for certain vowels) proportion scores to the matching face were entered into a between-subjects ANOVA that included gender (i.e., male or female), side of the match (i.e., left or right), heard-vowel (i.e., /i/ or /u/), and lip-shape (i.e., lip-spread or lip-round). A significant interaction between the heard-vowel and the lip-shape was found, $F(1, 48) = 4.60$, $p = .037$, $\eta^2 = .015$, but no other interactions or main effects reached significance. The simple main effect within the group of infants who heard /i/ sounds (and thus matched to the [i]-face) showed that gaze proportions were not different in the lip-spreading ($M = 47.0\%$; $SD = 26.3\%$) versus

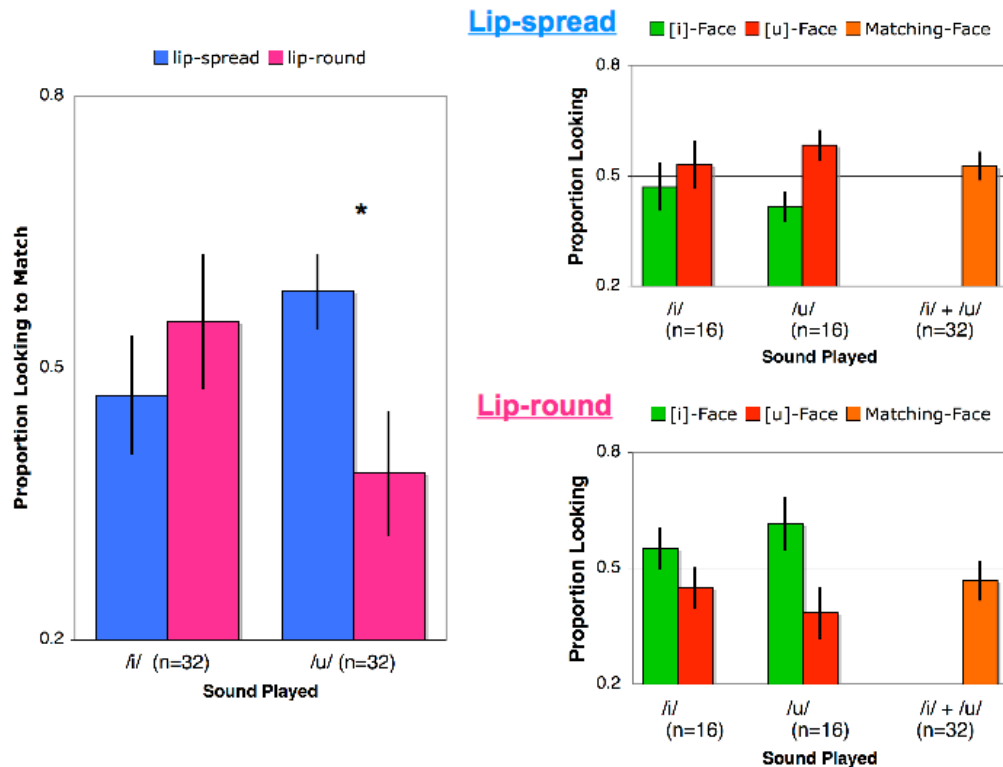
lip-rounding groups ($M = 55.1\%$; $SD = 30.0\%$), $F(1, 48)^8 = .77$, $p = .39$, $\eta^2 = .005$.

However, within the group of infants who heard /u/ sounds (and thus matched to the [u]-face), gaze proportions were significantly longer in the lip-spreading ($M = 58.4\%$; $SD = 16.7\%$) *versus* lip-rounding groups ($M = 38.4\%$; $SD = 27.5\%$), $F(1, 48) = 4.64$, $p = .036$, $\eta^2 = .031$ (see Figure 4.6).

⁸ Degrees of freedom are the same as the omnibus comparison, since the pooled estimate of the error was used in simple main effects analysis.

Figure 4.6 – Results showing the interaction between lip-shape and heard vowel (Experiment 3).

The left side of the figure illustrates the interaction between heard vowel and lip-shape; note that the y-axis on this graph is the proportion of gaze to the face that matches the sound played. Simple main effects analysis reveals a significant difference in matching proportions within the /u/-vowel group (i.e., the asterisk). The right side of the figure illustrates looking patterns within each of the lip-shape groups. Error bars indicate standard error.



Notice from the right-side Figure 4.6 that this interaction between lip-shape and heard-vowels has another conceptual interpretation. Among infants who were in the lip-spreading group, there was not a significant trend to look longer at the [u]-face ($M = 55.7\%$; $SD = 21.9\%$), $F(1, 48) = 1.52$, $p = .22$, $\eta^2 = .012$, but analysis of individual preferences revealed that 24 of 32 infants in this group preferred the [u]-face, which was significant by a binomial test, $p = .004$. Furthermore, this group showed no evidence of

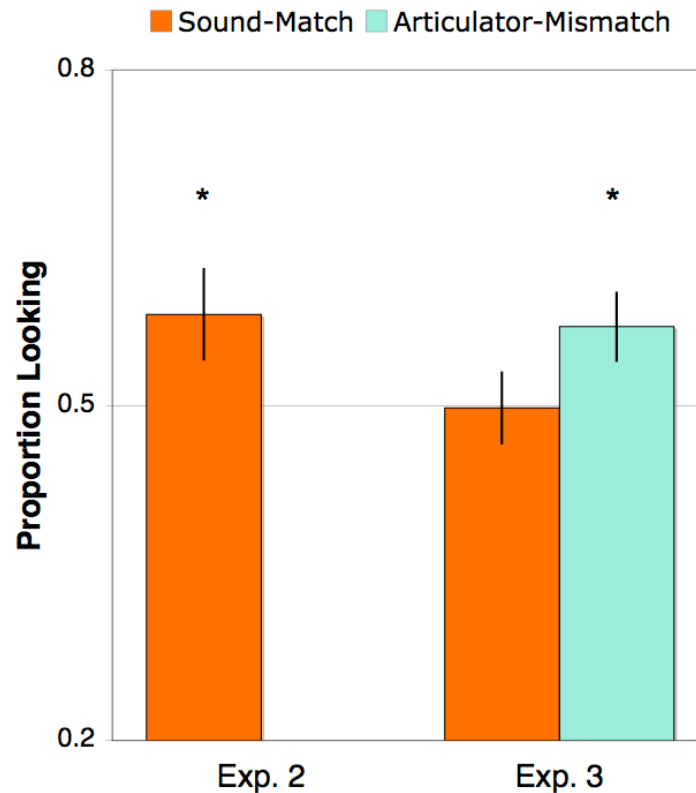
matching ($M = 52.7\%$; $SD = 22.4\%$), $F(1, 48) = .35$, $p = .56$, $\eta^2 = .002$ and only 18 of 32 infants looked longer at the matching face by a binomial test, $p = .30$.

Among infants who were in the lip-rounding group, the tendency to look more at the [i]-face approached significance ($M = 58.3\%$; $SD = 28.4\%$), $F(1, 48) = 3.23$, $p = .079$, $\eta^2 = .019$. Analysis of individual infants' face preferences also showed this pattern (although not significantly so), revealing that 20 of 32 infants showed this [i]-face preference, $p = .11$ by a binomial test. Furthermore, this group also showed no evidence of matching ($M = 46.8\%$; $SD = 29.5\%$), $F(1, 48) = .48$, $p = .49$, $\eta^2 = .004$, and exactly half of the 32 infants tested in this group looked longer at the matching face by a binomial test, $p = .57$. In summary, neither lip-shape group showed any evidence of cross-modal matching; rather the lip-spreading group showed some evidence of looking more at the [u]-face, while the lip-rounding group showed a trend to look more at the [i]-face.

These patterns suggest that infants gazed more at the face that articulated something different from what they were doing themselves. To investigate whether this general pattern held when looking at both lip-shape groups together, another analysis that explored the effects of lip-shape and matching behaviour was conducted. This analysis showed that infants looked 57.0% ($SD = 25.2\%$) of the time at the articulatory *mismatching* face, which was significantly greater than chance, $t(63) = 2.23$, $p = .029$, $d = .28$. In addition, 44 of 64 infants showed this pattern of preference, which was significant by a binomial test, $p = .002$. Thus, unlike in Experiment 2, where infants showed the classic preference for the face that matched the sound, infants in Experiment 3 preferred the face that was an articulatory mismatch with their self-produced lip-shapes (see Figure 4.7).

Figure 4.7 – A summary of infants' performance in Experiments 2 and 3.

Infants preferred the face that matched the heard vowel in Experiment 2. This was not the case in Experiment 3. Rather, infants there preferred the face that was a mismatch with their own articulatory configuration. Error bars indicate standard error.



4.4.3 Discussion

These results showed that lip-spreading when chewing on a large object, or lip-rounding when sucking on pacifier or fingertip can disrupt infants' ability to match /i/ and /u/ auditory sounds onto visual [i]- and [u]-faces, all of which contain related lip-shape information. These data are particularly exciting, because it is the first report showing that articulatory-motor behaviours can affect the perceptual processing of speech in pre-verbal infants. More specifically, these results show that articulatory information, even when is embedded in a non-speech gesture, can affect infants' ability to link related speech information across visual and auditory modalities.

Two more specific conclusions about these data can be drawn, which are related to the idea that articulatory mappings are used to link speech information across sensory modalities (Kent & Vorperian, 2007; Kuhl & Meltzoff, 1984, 1988). First, these data show that the lack of cross-modal matching was not the result of generalized interference: rather, infants' perceptual patterns were articulator-specific. Infants who were lip-spreading looked more at the [u]-face, while infants who were lip-rounding looked more at the [i]-face. Thus, a *contrast effect* was observed, such that infants looked more at the face that mismatched what they themselves were doing. This contrast effect was powerful enough to override the originally reported preference for the sound-based match in Experiment 2, directly implicating articulatory mappings in cross-modal speech perception.

Second, these results further suggest that the earliest connections between speech perception and articulatory processes are broadly tuned, sensitive to the articulator positions achieved during chewing, sucking, and otherwise holding objects in the mouth. Indeed, these are all explicitly non-speech movements: not only do these activities preclude vocalization, but the motor programs and muscle groups used to make these motions have unique coordinative profiles (Finan & Barlow, 1998; Steeve et al., 2008). Thus, by at least by 4.5 months of age, our results show that speech motor information used in an articulatory mapping is specified at a relatively low-level, where simple features or atomistic elements of articulation (i.e., lip-rounding, lip-spreading, jaw-opening, etc.) are linked to visual and auditory speech information. This not only provides strong evidence that speech perception and articulatory behaviour are linked from early in infancy, but further suggests that the basis of early speech motor control is broadly specified, with its origins in low-level motor information that is not specific to speech.

4.5 General Discussion

This study was designed to investigate the developmental origins of the links between speech perception and production, the basis of which may be mappings between articulatory, auditory, and visual speech information in infancy (Kent & Vorperian, 2007; Kuhl & Meltzoff, 1984, 1988). Three experiments investigated the role of articulatory mappings in cross-modal speech perception: first, 4.5-month-old infants' baseline preferences for looking at [i]- and [u]-faces were measured in Experiment 1. Second, infants' preference to look at cross-modally matching faces was replicated in Experiment 2. Third, it was shown in Experiment 3 that when infants make lip-shapes similar to [i] and [u] articulations, cross-modal matching is disrupted and an overall preference for the incongruent face is found (i.e., a contrast effect).

Interference and contrast effects in speech perception and production have not previously been reported, but they are far from unprecedented in other domains. For example, research in action perception shows that producing arm or hand movements can selectively interfere with the perceptual identification of visual stimuli that share features with the executed movement (James & Gauthier, 2009; Müsseler & Hommel, 1997). More explicit contrast effects have also been reported: certain actions (i.e., lifting a heavy weight or drawing a rising arc) while perceiving related visual displays (i.e., videos of weight-lifting or of a point-light moving in a rising arc) can bias perceptual identification away from percepts that share features with the performed actions (i.e., guessing that a lighter weight was lifted, or that the dot moved in a flatter arc) (Grosjean, Zwicker, & Prinz, 2009; Hamilton, Wolpert, & Frith, 2004; Zwicker, Grosjean, & Prinz, 2008; see Schütz-Bosbach & Prinz, 2007 for review). Theoretical models have explained these results by suggesting that both perceptual and motor processes share information that is mutually activated when actions are performed or when events are perceived.

This shared information, however, is withheld from perceptual analysis when simultaneously executing an action, resulting in perceptual interference or contrast effects associated with concurrent action production (Hamilton et al., 2004; Hommel et al., 2001; Prinz & Hommel, 2002).

A similar mechanism may underlie the contrast effects observed here. Engaging in articulatory movements while perceiving talking faces may have activated common features shared between different modal representations of speech. This mutual activation, rather than facilitating perception of the congruent cross-modal match, may instead have biased perceptual preferences away from stimuli (i.e., the visual faces) that contained similar articulatory information. Further research is required to determine if there are other situations in which contrast effects are seen, as the precise mechanisms that sometimes reveal contrast effects and sometimes reveal assimilation are not well described, even in the domain of action perception (Schütz-Bosbach & Prinz, 2007).

Recall, as well, that previous work in the adult speech perception literature has shown assimilation, rather than contrast effects: we speculate that this difference is related to the rapid development of speech-specific motor programs in infancy. For example, coordinative and dynamic structures in speech motor control already distinguish babbling from chewing and sucking by 9 months of age, and distinguish speech from other orofacial behaviours by 15 months of age (Moore & Ruark, 1996; Steeve et al., 2008). Future work may show that contrast effects are not stable across development, or even reappear as assimilation effects as infants gain more experience producing speech and as they develop more sophisticated and organized articulatory schemas.

Our work provides support for two broad-based conclusions concerning the developing link between speech perception and production. First, our results provide

direct evidence for the existence of articulatory mappings between auditory and visual speech information in 4.5-month-old infants. This is important, because it reveals linkages between articulatory behaviours and speech perception in infants who are not yet babbling, or producing clear speech. This provides striking evidence for an early basis of connections between speech perception and production, suggesting that such links are experience-independent. Second, our results further suggest that these articulatory mappings are specified at a level that is not limited to adult-like speech gestures. Rather, articulatory representations in 4.5-month-old infants appear to be specified broadly enough to allow lip-shapes achieved during non-speech movements to influence the audiovisual perception of speech. Together, these findings provide an exciting springboard from which one may further study how speech perception and production come to be linked from early in infancy, and how this link can inform separate literatures on speech perception and production.

4.6 References

- Baier, R., Idsardi, W. J., & Lidz, J. (2007). Two-month-olds are sensitive to lip rounding in dynamic and static speech events. In J. Vroomen, M. Swerts, & E. Krahmer (Eds.), *Proceedings of the International Conference on Auditory-Visual Speech Processing*. Retrieved from http://spitswww.uvt.nl/Fsw/Psychologie/AVSP2007/papers/BaierIL_AVSP2007.pdf
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange & J. J. Jenkins (Eds.), *Speech perception and linguistic experience: Issues in cross-language research*, Cross-language speech perception (pp. 171-204). Timonium, MD: York Press.
- de Boysson-Bardies, B., Halle, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16(1), 1-17. doi:10.1017/S0305000900013404
- de Boysson-Bardies, B., Sagart, L., & Durand, C. (1984). Discernible differences in the babbling of infants according to target language. *Journal of Child Language*, 11(1), 1-15.
- de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67(2), 297-319.
- 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
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19(5), 381-385. doi:10.1016/j.cub.2009.01.017
- Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech & Hearing Research*, 38(6), 1199-1211.
- Finan, D. S., & Barlow, S. M. (1998). Intrinsic dynamics and mechanosensory modulation of non-nutritive sucking in human infants. *Early Human Development*, 52(2), 181-197. doi:10.1016/S0378-3782(98)00029-2
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14(1), 3-28.
- Grosjean, M., Zwicker, J., & Prinz, W. (2009). Acting while perceiving: assimilation precedes contrast. *Psychological Research*, 73(1), 3-13.
- Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105(4), 611-633.
- Hamilton, A., Wolpert, D., & Frith, U. (2004). Your own action influences how you perceive another person's action. *Current Biology*, 14(6), 493-498. doi:10.1016/j.cub.2004.03.007

- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402. doi:10.1038/nrn2113
- Hommel, B., Bertoncini, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24(05), 849-878.
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1245-1248. doi:10.1073/pnas.0810063106
- James, K. H., & Gauthier, I. (2009). When writing impairs reading: Letter perception's susceptibility to motor interference. *Journal of Experimental Psychology: General*, 138(3), 416-431. doi:10.1037/a0015836
- Kent, R. D., & Vorperian, H. K. (2007). In the mouths of babes: Anatomic, motor, and sensory foundations of speech development in children. In R. Paul (Ed.), *Language disorders from a developmental perspective: Essays in honor of Robin S. Chapman*. (pp. 55-81). Mahwah, NJ: Lawrence Erlbaum.
- Kessen, W., Levine, J., & Wendrich, K. A. (1979). The imitation of pitch in infants. *Infant Behavior and Development*, 2, 93-99.
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577), 1138-1141.
- Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior & Development*, 7(3), 361-381.
- Kuhl, P. K., & Meltzoff, A. N. (1988). Speech as an intermodal object of perception. In A. Yonas (Ed.), *Perceptual development in infancy*, The Minnesota Symposium on Child Development. (Vol. 20, pp. 235-266). Hillsdale, NJ: Lawrence Erlbaum.
- 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.
- 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.
- Legerstee, M. (1990). Infant use of multimodal information to imitate speech sounds. *Infant Behavior and Development*, 13(3), 343-354.
- Levitt, A. G., & Utman, J. A. (1992). From babbling towards the sound systems of English and French: A longitudinal two-case study. *Journal of Child Language*, 19(1), 19-49. doi:10.1017/S0305000900013611
- Lewkowicz, D. J. (2000). Infants' perception of the audible, visible, and bimodal attributes of multimodal syllables. *Child Development*, 71(5), 1241-1257. doi:10.1111/1467-8624.00226
- Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, 46(1), 66-77. doi:10.1037/a0015579

- 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
- Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Publishers.
- MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science*, 219(4590), 1347-1349. doi:10.1126/science.6828865
- MacNeilage, P. F., & Davis, B. L. (1993). Motor explanations of babbling and early speech patterns. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. F. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life*, NATO Science Series (Vol. 69, pp. 341–352). Norwell, MA: Kluwer.
- 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.2009.09.064
- McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research*, 44(3), 670-684. doi:10.1044/1092-4388(2001/054)
- Moore, C. A., & Ruark, J. L. (1996). Does speech emerge from earlier appearing oral motor behaviors? *Journal of Speech and Hearing Research*, 39(5), 1034-1047.
- Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, 29(31), 9819-9825. doi:10.1523/JNEUROSCI.6018-08.2009
- Müsseler, J., & Hommel, B. (1997). Blindness to response-compatible stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 861-872. doi:10.1037/0096-1523.23.3.861
- Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), 20470.
- Nazzi, T., Bertoncini, J., & Bijeljac-Babic, R. (2009). A perceptual equivalent of the labial-coronal effect in the first year of life. *The Journal of the Acoustical Society of America*, 126, 1440.
- Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komishan, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology, Vol. 1: Production* (pp. 92–112). New York: Academic Press.
- Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior & 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.

- Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10598-10602.
- Prinz, W., & Hommel, B. (2002). *Common mechanisms in perception and action*. New York: Oxford University Press.
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351-360. doi:10.1038/nrn2811
- Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, 21(5), 188-194. doi:10.1016/S0166-2236(98)01260-0
- Ruzza, B., Rocca, F., Boero, D. L., & Lenti, C. (2006). Investigating the musical qualities of early infant sounds. *Annals of the New York Academy of Sciences*, 999, 527–529.
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, 23(2-3), 429-435.
- Schütz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: action-induced modulation of perception. *Trends in Cognitive Sciences*, 11(8), 349-355. doi:10.1016/j.tics.2007.06.005
- Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2006). Lending a helping hand to hearing: another motor theory of speech perception. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 250 - 285). Cambridge, MA: Cambridge University Press.
- Sloutsky, V. M., & Robinson, C. W. (2008). The Role of Words and Sounds in Infants' Visual Processing: From Overshadowing to Attentional Tuning. *Cognitive Science*, 32(2), 342-365. doi:10.1080/03640210701863495
- Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komishan, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology: Vol 1., Production* (Vol. 1, pp. 73-92). New York: Academic Press.
- Steeve, R. W., Moore, C. A., Green, J. R., Reilly, K. I., & McMurtrey, J. R. (2008). Babbling, chewing, and sucking: Oromandibular coordination at 9 months. *Journal of Speech, Language, and Hearing Research*, 51(6), 1390-1404. doi:10.1044/1092-4388(2008/07-0046)
- Vihman, M. M. (1991). Ontogeny of phonetic gestures: Speech production. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception: Proceedings of a conference to honor Alvin M. Liberman*. Mahwah, NJ: Lawrence Erlbaum.
- Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21(1/2), 61-82.
- Walton, G. E., & Bower, T. G. R. (1993). Amodal representation of speech in infants. *Infant Behavior and Development*, 16(2), 233–243.

- Werker, J. F. (1993). The contribution of the relation between vocal production and perception to a developing phonological system. *Journal of Phonetics*, Phonetic development, 21(1-2), 177-180.
- Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, 50, 509-535.
- Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, 46(3), 233-251. doi:10.1002/dev.20060
- Whalen, D. H., Levitt, A. G., & Goldstein, L. M. (2007). VOT in the babbling of French- and English-learning infants. *Journal of Phonetics*, 35(3), 341-352. doi:10.1016/j.wocn.2006.10.001
- Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 107(2), 592-597. doi:10.1073/pnas.0904774107
- Zwicker, J., Grosjean, M., & Prinz, W. (2008). A contrast effect between the concurrent production and perception of movement directions. *Visual Cognition*, 16(7), 953-978. doi:10.1080/13506280701653586

5: GENERAL DISCUSSION

5.1 Executive summary of empirical chapters

One reason the study of perception is so fascinating is because it involves much more than simply recording or translating real-world events into sensory signals. In the study of perception, understanding how the brain analyzes these sensory signals is just as important (if not more so) than the sensory signal itself. In the literature on embodied perception, motor processes play a particularly important role in this perceptual analysis (Knoblich, 2008; Prinz & Hommel, 2002; Pulvermüller & Fadiga, 2010). As a case study of embodied perception, this dissertation focuses on the specific role that articulatory-motor processes have in perceptual analysis of the speech signal. Specifically, two empirical chapters investigate origins in (processing) of articulatory-motor influences on speech perception, while a third investigates the origin of links between speech perception and production in development.

In terms of processing, an emerging body of empirical research in the last 5 years has reported that articulatory processes are related or linked to speech perception (D'Ausilio et al., 2009; Meister, S. M. Wilson, Deblieck, Wu, & Iacoboni, 2007; Möttönen & Watkins, 2009; Nasir & Ostry, 2009; Roelofs, Özdemir, & Levelt, 2007; Sams, Möttönen, & Sihvonen, 2005). However, the precise source of perceptual modulation in these reports has been a topic of some controversy: some have suggested that these influences on perception stem not from motor or sensorimotor information resulting from articulation, *per se*, but rather from corollary activation of high-level conceptual categories or auditory abstractions (i.e., phonological representations), which commonly accompany articulation (Hickok, Holt, & Lotto, 2009; Mahon & Caramazza, 2008). This

claim is tested in Chapter 2, the results of which suggest that information about the movements of articulators themselves can indeed be one source of perceptual modulation, and one that is independent from corollary activation of symbolic or auditory representations. This is important in establishing that the perceptual modulation observed in the existing literature is indeed motor-related. Moreover, results from Experiment 2 in Chapter 2 further suggest that perceptual modulation from these sources is comparable in size to modulation that is derived from auditory influences, weakening theoretical arguments against embodiment.

Chapter 3 further explores the processing origins of articulatory-motor influences in speech perception, but asks instead how different levels of organization within speech motor control might variously contribute to perceptual modulation. Previous motor theories of speech perception, for example, had always assumed that relatively abstract task-level information about the coordinative dynamics of speech motor control are the appropriate level of analysis, and that this is the level at which sensorimotor information is specified in the perceptual processing of speech (Fowler, 1996; Fowler & Galantucci, 2005; Galantucci, Fowler, & Turvey, 2006). However, results from Chapter 3 suggest something different: low-level articulatory information, even that which is part of an explicitly non-speech gesture, can also exert some influence on speech perception. In the experiment reported here, participants maintained a single articulatory configuration while breath holding, which was very different from the type of dynamic, speech-like movements studied in Chapter 2. Nevertheless, there was still a significant impact on speech perception when maintaining these articulatory positions. Thus, a second basic claim made in this dissertation suggests that information at different levels in the hierarchy of speech motor control can be a source of perceptual influence.

These first two chapters pave the way to further ask about the developmental origins of articulatory-motor influences in speech perception (i.e., Chapter 4). As very young infants do not yet produce the dynamically coordinated articulatory movements typically elicited in adult studies, Chapters 2 and 3 allowed for the possibility that simple, non-speech articulatory configurations could still shape infants' perceptual preferences for speech. Results confirmed this hypothesis, showing that performance in a cross-modal matching task with talking faces was disrupted when 4.5-month-old infants achieved certain articulatory configurations (i.e., lip-spreading or lip-rounding) relevant to the presented speech stimuli (i.e., /i/ vowels and [i]-faces containing lip-spreading versus /u/ vowels and [u]-faces containing lip-rounding). Moreover, results from Chapter 4 suggested that achieving these articulatory configurations resulted in more specific patterns than simple disruption of matching. Rather, infants who achieved a particular lip-shape showed a selective bias away from visual speech information that was compatible with their own articulations. This suggests that low-level elements of articulatory behaviour, even when embedded in non-speech movements like sucking and chewing, can exert an articulator-specific influence on perceptual processing of speech. Results also suggest that the developmental link between speech perception and production is already in place at a developmentally early stage, even before infants begin babbling or producing well-formed vowels.

5.2 Implications and directions for future research

Together, these empirical chapters provide the basis for the two broad characterizations of perception and perceptual development outlined in the introductory chapter: this dissertation suggests that mental processes involved in speech are in some sense *embodied* rather than purely *symbolic*, and that the course of development in the coupling between speech perception and production is one of *differentiation* rather than

of *enrichment*. Implications of these data for both embodiment and differentiation approaches are discussed below, as are directions for future research that are related to these implications. Finally, an additional issue is raised: what does the *direction* of the articulatory-motor influence in Chapter 4 (i.e., a contrast effect), compared to the directions of influence in Chapters 2 and 3 (i.e., assimilation effects) reveal about perceptual processing?

5.2.1 Theoretical implications for embodied approaches

Each empirical chapter provides evidence that information about the movements or positions of the articulators can be a source of perceptual modulation in speech. While Chapter 2 is important in arguing against alternative explanations of embodied phenomena and Chapter 4 presents evidence relevant for understanding the development of embodied processes, the results most relevant for advancing theories of adult perception that already take an embodied approach come from Chapter 3. This chapter suggest that low-level articulatory information achieved in the context of explicitly non-speech gestures can also influence speech perception, including information from maintaining static articulatory positions. As mentioned above, theories in speech research have not previously considered the possibility that this sort of articulator-level information plays a significant role in speech representations. Rather, it has always been assumed that more high-level information about the coordinative dynamics of all the articulators is the locus of information that is relevant to speech (Fowler & Galantucci, 2005; Galantucci et al., 2006). This assumption is based, in turn, on the prodigious literature in speech motor control that suggests there is hierarchical organization in the motor systems involved in speech production (Gracco, 1994; Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984; Saltzman & Munhall, 1989; Shaiman &

Gracco, 2002; Sternberg, Knoll, Monsell, & Wright, 1988; Tremblay, Houle, & Ostry, 2008).

Our current results suggest that the hierarchical organization established in speech motor control cannot be assumed in the study of links between speech perception and production. Rather, there is likely a more subtle relation between levels of abstraction in speech production and in speech perception, some hints about which are provided by our results. Recall that Ito et al. (2009) showed that deforming facial skin and muscles in ways that were similar to the dynamic speech-like movements associated with certain vowels could affect the categorical identification of those vowels, but that neither static positions nor dynamic movements delivered at faster rate than speech (i.e., “twitches”) had a similar effect. The data from Chapter 3 suggest that Ito et al. (2009) underestimated the influence of simply maintaining static articulatory positions, but their results in combination with those from Chapter 3 are still compatible with the idea that dynamic speech-like movements exert a *stronger* effect on perception than static ones. Specifically, Chapter 3 showed only that static articulatory positions could influence reaction times (RTs), rather than both RTs and categorical identifications. This suggests, in turn, that if articulatory-motor information more closely approximates actual speech gestures, then more pronounced effects on speech perception should be observed.

Interestingly, this discussion in the speech domain parallels a similar debate in the action-perception literature. Common-coding theories disagree with theories based on forward models about whether information shared between action and perception is categorical, or more graded in nature (Hamilton, Wolpert, & Frith, 2004; Schubö, Aschersleben, & Prinz, 2001; Zwicker, Grosjean, & Prinz, 2010). The categorical view that is held among common-coding theories is supported by data, which show that the

influence of performed actions on perceptual judgements of a visual event is not affected by the degree of overlap between the two (Zwicker et al., 2010). Theories invoking forward models, however, predict that the outputs of motor simulations used in perceptual analysis are graded in nature, and the degree of feature overlap between an action and a perceptual event influences the degree of perceptual influence in a more continuous fashion (Hamilton et al., 2004; Haruno, Wolpert, & Kawato, 2001; Wolpert & Kawato, 1998).

Our results preliminarily suggest that the degree of similarity that articulatory motions have to actual speech gestures may influence speech perception in a graded fashion. In other words, the information shared between articulatory-motor and perceptual processes are likely graded in nature, and not categorical. However, this is one area that will require further research: for example, the precise dimensions along which articulatory-motor information and speech perception may be classified as being more “similar” or “dissimilar” remain unknown. Are some distinctions (i.e., dynamic movements or static positions; place or manner of articulation; etc.) more important in determining the degree of overlap than others?

5.2.2 Theoretical implications for differentiation approaches

Chapter 4 presents some of the first data to examine the influence of infants’ articulatory behaviour on speech perception, showing that articulatory-motor information embedded in non-speech gestures can still influence 4.5-month-old infants’ behaviour in an audiovisual speech perception task. This provides direct evidence that speech information in both auditory and visual modalities is commonly mapped onto an articulatory format (Kent & Vorperian, 2007; Kuhl & Meltzoff, 1984, 1988). What, then, is the developmental origin of these articulatory mappings?

On the one hand, several studies have suggested that these mappings are learned from experience producing vocalizations, even at this young age. For example, infants' vocalizations become increasingly vowel-like between 3 and 4 months of age, when this cross-modal matching behaviour is typically observed (Kuhl & Meltzoff, 1996). Recent evidence, however, has suggested that cross-modal matching is robust in infants as young as 2 months of age (Patterson & Werker, 2003), which weakens, but does not invalidate this experience-based hypothesis. Other evidence for this hypothesis comes from a study reporting an asymmetry between certain types of cross-modal matches: infants at 4 months of age can cross-modally match videos and sounds of bilabial trills (i.e., flapping the lips together in a quick, trill-like motion), but not videos and sounds of whistling. Crucially, infants are commonly assumed to have experience producing trills, but are unlikely to have experience producing whistles (Mugitani, Kobayashi, & Hiraki, 2008).

On the other hand, several lines of evidence suggest that articulatory mappings are established by birth. Many studies have shown, for example, that neonates only a few days old can imitate facial expressions modelled by adults (Meltzoff & M. K. Moore, 1977, 1983, 1989). Recent evidence has further suggested that neonates engage in vocal imitation, suggesting links between speech perception and motor processes at this stage. When listening to either /a/ or /m/ sounds modelled by a visible experimenter, neonates will preferentially make congruent mouth-shapes by opening or closing their mouth. This occurs, however, even if neonates spend a substantial amount of time in the experiment with their eyes closed (Chen, Striano, & Rakoczy, 2004). This evidence is suggestive of the idea that links between speech perception and production, as well as

the general phenomenon of imitation⁹, both have their roots in mappings between vision, audition, and the motor system that are present at birth.

Whatever the status of the initial state, it remains likely that the characteristics of articulatory mappings are refined by both experience hearing and producing speech, and by the rapidly progressing physiological changes in the vocal tract of young infants throughout infancy and early childhood (Vorperian et al., 2005). For example, data from Chapter 4 show that articulatory mappings in infancy are broadly specified, and not specific to speech-like gestures. Data from adults presented in Chapters 3 and imply that the power of these non-speech articulatory effects are not as effective in adulthood. For example, in Chapter 4, articulatory mappings that were based on dynamic and speech-like movements influenced perceptual identity, but, in Chapter 3, articulatory information in non-speech gestures only influenced reactions times in perceptual identification tasks, having a necessarily subtler effect on perception. This pattern of results implies a developmental trajectory similar in spirit to differentiation accounts offered by direct-realist theories of speech perception, which have always assumed that gestural representations underlying speech perception become more detailed and language-specific in development with more native language input (Best, 1995). The data from this dissertation go beyond this view, however, providing specific evidence for the link between information in infants' actual orofacial movements and their early perceptual preferences for speech, and charting this influence in adulthood. In summary, while it is likely that there are mappings between sensorimotor and perceptual systems from the beginnings of life, it is also likely that these mappings are broadly specified, and not specific to speech gestures until some later point in development.

⁹ It seems worth mentioning that the term "imitation," carries with it a number of other connotations: for example that it is dependent on social factors and provides an early basis for social knowledge (Meltzoff & M. K. Moore, 1992, 1997). The phenomenon of imitation is likely independent of the effects reported here, which is supported by the fact that we report a *contrast* effect, which is not predicted by any existing accounts of imitation.

Future research will be needed to obtain more precise measurements of infants' experience producing vocalizations, and the relation of this experience with infants' developing speech perception abilities. Current evidence, however, shows some striking parallels in the developmental patterns found within infant studies of speech perception and within similar studies on speech motor control. On the perception side, developmental processes follow a general pattern of differentiation, described more precisely as reorganization of phonetic perception: perception is initially broad, while sensitivity to native-language contrasts improves over several years and sensitivity to non-native contrasts declines by at least 12 months of age (Best, 1993; Gervain & Werker, 2008; Werker, 1995). A similar pattern of development is seen in infants' perception of the native and non-native language rhythms in the visual modality (i.e., detecting linguistic information from talking faces) (Weikum et al., 2007). Several studies have further suggested different mechanisms by which native-language input drives these language-specific changes in speech perception, including the influences of statistical information (Maye, Werker, & Gerken, 2002; Werker et al., 2007; Yoshida, Pons, Maye, & Werker, 2010), social interaction (Kuhl, Tsao, & Liu, 2003), and the co-occurrence of speech events with other cues in the environment (Yeung & Werker, 2009). Recent work has also shown that the effects of experience may also be gated by some language-independent (i.e., maturational) factors (Peña, Pittaluga, & Mehler, 2010).

On the production side, strikingly similar patterns of development are seen, although somewhat delayed compared to perception. For example, speech motor control is rapidly differentiated from articulatory-motor processes that are related to other non-speech behaviours (E. M. Wilson, Green, Yunusova, & C. A. Moore, 2008). This is a process that has begun by at least 9 months of age, when babbling behaviours show

basic differences in motor coordination from sucking and chewing (Steeve, C. A. Moore, Green, Reilly, & McMurtrey, 2008). Just as in the speech perception literature, speech motor control continues to specialize and become more efficient through the first few years of life (Smith, 2006), while alimentary behaviours (i.e., movements of the articulators related to nursing and eating) continue to be further differentiated from these speech movements (C. A. Moore & Ruark, 1996; E. M. Wilson et al., 2008). Again, as in the perception literature, exposure to the native-language has also been found to influence the characteristics of babbling and early word production (de Boysson-Bardies, Halle, Sagart, & Durand, 1989; de Boysson-Bardies & Vihman, 1991; McCune & Vihman, 2001; Whalen, Levitt, & L. M. Goldstein, 2007), and the quality of these vocalizations is further guided by social interactions with caregivers (M. H. Goldstein & Schwade, 2008), which are mediated by feedback that co-occurs with infants' production of vocal events (Gros-Louis, West, M. H. Goldstein, & King, 2006). Again, as in speech perception, these experience-based influences on infant speech production may still be gated by maturational factors common across several domains in the motor system (Dolata, Davis, & MacNeilage, 2008; Fagan, 2009; Iverson & Fagan, 2004).

The fact that both the speech perception and speech motor control literatures show patterns of differentiation suggests that the link between perception and articulatory-motor processes also follows a similar pattern. In other words, links between speech perception and production are likely specified quite broadly (i.e., not specific to speech gestures, *per se*) in early stages of development, but are refined and shaped by the experience that infants have producing babbles and other speech-like vocalizations in their first and second years of life. This would eventually result in the kind of strong connections between speech-like gestures and perception (e.g., those effects observed in Chapter 4), while vestigial links between low-level articulatory information and speech

perception decline (although perhaps still detectable, as shown in Chapter 3). Evidence for this hypothesis is provided in Chapter 4, which suggests that articulatory information can be specified at a relatively broad level when influencing perception. However, a substantial amount of future research is needed to confirm this hypothesis, and to further understand how the speech motor processes that provide the substrate for the articulatory mappings studied in Chapter 4 change as a function of rapid development in both the speech perception and production systems within the first few years of life.

5.2.3 Contrast *versus* assimilation effects

A final unresolved issue is the discrepant directions of articulatory-motor influence observed in the empirical chapters of this dissertation. Chapter 4 reports a contrast effect between articulatory behaviour and infants' perceptual preferences (i.e., perceptual processing is biased away from articulatory information), while Chapters 2 and 3 as well as previous studies in the adult literature examining perception-production couplings in speech have found assimilation effects (i.e., where perceptual processing is biased towards articulatory information) (e.g., D'Ausilio et al., 2009; Ito et al., 2009; Möttönen & Watkins, 2009; Nasir & Ostry, 2009; Sams et al., 2005). How might these differences in the direction of articulatory-motor influences inform the study of perceptual analysis?

Within the action-perception literature, both contrast effects (Hamilton et al., 2004; Müsseler & Hommel, 1997; Zwickel, Grosjean, & Prinz, 2008, 2010) and assimilation effects (Craighero, Fadiga, Rizzolatti, & Umiltà, 1999; Lindemann & Bekkering, 2009; Repp & Knoblich, 2007, 2009; Tucker & Ellis, 1998; Wohlschläger, 2000; Wohlschläger & Wohlschläger, 1998) have previously been found. Some have suggested that the differences in the “functional relatedness” between actions and perceptual events are useful in predicting the direction of perceptual influences (Zwickel

et al., 2010). On the one hand, paradigms in which motor behaviours are functionally unrelated to the perceptual task tend to show contrast effects (Hamilton et al., 2004; Müsseler & Hommel, 1997; Zwickel et al., 2008, 2010). For example, the amount of weight lifted by an actor in a video is judged to be heavier than it actually is when observers making these judgments simultaneously lift relatively lighter weights themselves (Hamilton et al., 2004). This contrast effect may be stem from the fact that participants knew that the motor activity was independent from (i.e., functionally unrelated to) the perceptual task. On the other hand, studies which show assimilation effects have commonly established a functional relation between perceptual cues and performed actions. For example, several stimulus-response studies have shown that if a motor response (i.e., a grasping or rotating hand movement) is congruent with a go/no-go cue (i.e., objects that afford grasping, or flickered images that imply visual rotation), then faster response times are observed (Craighero et al., 1999; Lindemann & Bekkering, 2009; Tucker & Ellis, 1998). In other words, faster visual processing of the go/no-go cue (i.e., an assimilation effect) may have been observed because the cue itself is functionally related to the motor response.

It is perhaps tempting to use this concept of functional relatedness within the action-perception literature to help explain the discrepant data patterns from Chapters 2 and 3 on the one hand, and Chapter 4 on the other. Adults in Chapters 2 and 3 have likely had prodigious amounts of experience perceiving self-produced speech, and thus articulatory movements might be considered more “functionally related” to speech stimuli than for the pre-verbal infants tested in Chapter 4. Infants this age, of course, do not have comparable amounts of experience producing speech, and articulating while perceiving audiovisual speech may have been more akin to tasks in the action-perception literature, where the motor task is not seen as being related to the perception

task. This would predict assimilation effects for all adult experiments in the speech domain generally, while a contrast effect would be predicted in paradigms examining pre-verbal infants, like those tested in Chapter 4. Moreover, this hypothesis suggests also that the same infant paradigm described in this dissertation would result in assimilation effects at older ages, when infants might have more top-down expectations of how their articulatory movements are functionally related to the speech signal.

This explanation is not without problems, however. First, the precise definition of what makes a motor movement and a perceptual stimulus functionally related is not precisely outlined. Zwickel et al. (2010), who originally advocated this distinction, simultaneously classify go/no-go tasks that involve grasping (Craighero et al., 1999; Tucker & Ellis, 1998) as functionally relating perceptual processing and motor responses, but tasks where one must presses a key on a keyboard to elicit an auditory stimulus as functionally unrelated (Repp & Knoblich, 2007, 2009). Moreover, this idea of functional relatedness in the speech domain is equally unclear: are motor tasks where articulatory information is embedded in a non-speech gesture defined as functionally related or unrelated to speech processing (i.e., as in Chapter 3)? On its face, one might consider such articulatory information to be functionally unrelated to speech, precisely because these movements are not part of the dynamic coordination that is characteristic of speech motor control. If this were the case, however, this presents a problem for the functional relatedness account, since assimilation effects were found in both Chapters 2 *and* 3, which provided articulatory information in both dynamic speech-like contexts *and* static non-speech contexts.

A second problem is that the issue of contrast versus assimilation, even within the action-perception literature, cannot be completely reduced to task differences (see Schütz-Bosbach & Prinz, 2007 for review). Specifically, assimilation effects are

sometimes found in paradigms where the motor task is considered functionally irrelevant to the perceptual task, which is not a prediction made by the functional relatedness account. For example, making functionally unrelated rotational hand movements while identifying ambiguous visual displays as moving in either clockwise or counter-clockwise directions will bias perception of apparent motion in the same direction as the executed rotational movement (Wohlschläger, 2000). This differs from similar experiments, where participants also make simultaneous hand movements while perceiving visual displays, but nevertheless find contrast effects (Schubö et al., 2001). One possibility is that the degree of perceptual ambiguity may also play a role (i.e., in addition to functional relatedness): ambiguous perceptual stimuli used in a functionally unrelated task may still result in assimilation, rather than contrast (Zwicker et al., 2010). However, this distinction between ambiguous versus unambiguous stimuli still does not translate well into the speech domain, as assimilation effects have been found in adult paradigms that use speech embedded in noise, synthesized speech, or ambiguous speech continua (D'Ausilio et al., 2009; Ito et al., 2009; Möttönen & Watkins, 2009; Nasir & Ostry, 2009), as well as in paradigms that use naturally produced tokens of speech, which are relatively less ambiguous (Roelofs et al., 2007; Sams et al., 2005; also Chapter 2).

In summary, the distinction between contrast and assimilation effects is a controversial one in the action-perception literature (Schütz-Bosbach & Prinz, 2007), and further work will be needed to establish why sometimes one direction of influence is found, and sometimes the other. This debate has not previously been an issue in the speech domain, as the vast majority of studies investigating links between speech perception and production have shown assimilation effects. However, the data presented in Chapter 4 buck this trend, offering one of the first cases of a contrast effect in this literature. While the precise factors contributing to contrast effects in this study are

still unclear, it is likely not a coincidence that this study was also one of the first to test a population who does not have a significant amount of experience producing speech: pre-verbal infants. Further research will be needed to clarify what precise mechanisms are responsible for infants' unique patterns of behaviour, and how this may change through development as they acquire more experience producing and perceiving speech.

5.3 Conclusions and broader impacts

This dissertation offers three broad lessons for future research within the literature on embodiment and the study of perception in general. First, Chapter 2 reinforces the idea that a single motor process (e.g., articulating a syllable) can trigger a cascade of neural events, many of which can possibly influence perceptual analysis. Future research will need to be careful in parsing the influences from these different pathways by using more subtle and novel experimental designs. Second, as discussed in Chapter 3, very little attention has been devoted to understanding how the hierarchical structure of motor control can variously influence perceptual processes. Future behavioural and neurophysiological work in all perceptual domains must be careful in examining how these different levels of information may variously influence perception processes. Third, Chapter 4 reinforces the value of a developmental approach in advancing research questions in this field. Given how difficult it is to describe the complex and multi-layered interactions between perception and action in adults, developmental researchers are in the unique position of asking how these links are structured in individuals who do not have the same richly elaborated representations and well-developed processes in both perception and in motor control. The present data suggest, nevertheless, that some forms of perceptual-motor linkages, at least in the speech domain, are present from early in infancy and unlikely to be derived from

experience perceiving the products of one's own actions. Future research will need to consider how the structure of these early perception-production linkages changes as infants accrue more experience acting, and perceiving the results of these actions. This provides a jumping-off point from which embodiment researchers who are interested in more developed or mature processes in perception might begin.

Finally, it is anticipated that this research will eventually have implications for clinical practice, particularly in the study of developmental disorders in speech and language. A broader review of this literature is beyond the scope of the current research, but suffice it to say that most clinical research in this area has focused on the ways in which children have problems producing speech, rather than how these processes are linked, in turn, to perception. However, due in large part to influence from the mirror neuron literature (Rizzolatti & Craighero, 2004), recent studies in this area have begun to examine the possibility that disorders in perception are closely linked to disorders in production, perhaps causally so (Nijland, 2009). It is hoped that this dissertation research will eventually be useful in describing these links between perception and production in clinical practice as well, yielding more successful interventions and treatments for those individuals affected.

5.4 References

- Best, C. T. (1993). Emergence of language-specific constraints in perception of non-native speech: A window on early phonological development. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. F. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life*, NATO Science Series (Vol. 69, pp. 289-304). Norwell, MA: Kluwer.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange & J. J. Jenkins (Eds.), *Speech perception and linguistic experience: Issues in cross-language research*, Cross-language speech perception (pp. 171-204). Timonium, MD: York Press.
- de Boysson-Bardies, B., Halle, P., Sagart, L., & Durand, C. (1989). A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16(1), 1-17. doi:10.1017/S0305000900013404
- de Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67(2), 297-319.
- 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
- Craighero, L., Fadiga, L., Rizzolatti, G., & Umiltà, C. (1999). Action for perception: A motor-visual attentional effect. *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1673–1692.
- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, 19(5), 381-385. doi:10.1016/j.cub.2009.01.017
- Dolata, J. K., Davis, B. L., & MacNeilage, P. F. (2008). Characteristics of the rhythmic organization of vocal babbling: Implications for an amodal linguistic rhythm. *Infant Behavior & Development*, 31(3), 422-431. doi:10.1016/j.infbeh.2007.12.014
- Fagan, M. K. (2009). Mean Length of Utterance before words and grammar: Longitudinal trends and developmental implications of infant vocalizations. *Journal of Child Language*, 36(03), 495–527.
- Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, 99(3), 1730-41. doi:8819862
- Fowler, C. A., & Galantucci, B. (2005). The relation of speech perception and speech production. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 633-652). Hoboken, NJ: Wiley.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361-377.
- Gervain, J., & Werker, J. F. (2008). How infant speech perception contributes to language acquisition. *Language and Linguistics Compass*, 2(6), 1149–1170.

- Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19(5), 515-523. doi:10.1111/j.1467-9280.2008.02117.x
- Gracco, V. L. (1994). Some organizational characteristics of speech movement control. *Journal of Speech & Hearing Research*, 37(1), 4. doi:Article
- Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, 30(6), 509-516. doi:10.1177/0165025406071914
- Hamilton, A., Wolpert, D., & Frith, U. (2004). Your own action influences how you perceive another person's action. *Current Biology*, 14(6), 493-498. doi:10.1016/j.cub.2004.03.007
- Haruno, M., Wolpert, D. M., & Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Computation*, 13(10), 2201-2220.
- Hickok, G., Holt, L. L., & Lotto, A. J. (2009). Response to Wilson: What does motor cortex contribute to speech perception? *Trends in Cognitive Sciences*, 13(8), 330-331. doi:10.1016/j.tics.2009.05.002
- Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1245-1248. doi:10.1073/pnas.0810063106
- Iverson, J. M., & Fagan, M. K. (2004). Infant vocal-motor coordination: Precursor to the gesture-speech system? *Child Development*, 75(4), 1053-1066.
- Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, 10(6), 812-832. doi:10.1037/0096-1523.10.6.812
- Kent, R. D., & Vorperian, H. K. (2007). In the mouths of babes: Anatomic, motor, and sensory foundations of speech development in children. In R. Paul (Ed.), *Language disorders from a developmental perspective: Essays in honor of Robin S. Chapman*. (pp. 55-81). Mahwah, NJ: Lawrence Erlbaum.
- Knoblich, G. (2008). Motor contributions to action perception. In R. Klatzky, B. MacWhinney, & M. Berhmann (Eds.), *Embodiment, Ego-Space, and Action* (pp. 45-78). New York: Psychology Press.
- Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior & Development*, 7(3), 361-381.
- Kuhl, P. K., & Meltzoff, A. N. (1988). Speech as an intermodal object of perception. In A. Yonas (Ed.), *Perceptual development in infancy*, The Minnesota Symposium on Child Development. (Vol. 20, pp. 235-266). Hillsdale, NJ: Lawrence Erlbaum.
- 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.

- Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15), 9096-9101.
- Lindemann, O., & Bekkering, H. (2009). Object manipulation and motion perception: Evidence of an influence of action planning on visual processing. *Journal of Experimental Psychology: Human Perception and Performance*, 35(4), 1062-1071. doi:10.1037/a0015023
- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, 102(1-3), 59-70. doi:10.1016/j.jphysparis.2008.03.004
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101-B111.
- McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research*, 44(3), 670-684. doi:10.1044/1092-4388(2001/054)
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17(19), 1692-1696. doi:10.1016/j.cub.2007.08.064
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198(4312), 75-78. doi:10.1126/science.198.4312.75
- Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, 54(3), 702-709. doi:10.2307/1130058
- 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), 954-962. doi:10.1037/0012-1649.25.6.954
- Meltzoff, A. N., & Moore, M. K. (1992). Early imitation within a functional framework: The importance of person identity, movement, and development. *Infant Behavior & Development*, 15(4), 479-505. doi:10.1016/0163-6383(92)80015-M
- Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development & Parenting, Perceptual Development*, 6(3), 179-192.
- Moore, C. A., & Ruark, J. L. (1996). Does speech emerge from earlier appearing oral motor behaviors? *Journal of Speech and Hearing Research*, 39(5), 1034-1047.
- Möttönen, R., & Watkins, K. E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *Journal of Neuroscience*, 29(31), 9819-9825. doi:10.1523/JNEUROSCI.6018-08.2009
- Mugitani, R., Kobayashi, T., & Hiraki, K. (2008). Audiovisual matching of lips and non-canonical sounds in 8-month-old infants. *Infant Behavior & Development*, 31(2), 307-310.

- Müsseler, J., & Hommel, B. (1997). Blindness to response-compatible stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 861-872. doi:10.1037/0096-1523.23.3.861
- Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), 20470.
- Nijland, L. (2009). Speech perception in children with speech output disorders. *Clinical Linguistics & Phonetics*, 23(3), 222–239.
- 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., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8), 3823.
- Prinz, W., & Hommel, B. (2002). *Common mechanisms in perception and action*. New York: Oxford University Press.
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351-360. doi:10.1038/nrn2811
- Repp, B. H., & Knoblich, G. (2007). Action can affect auditory perception. *Psychological Science*, 18(1), 6–7.
- Repp, B. H., & Knoblich, G. (2009). Performed or observed keyboard actions affect pianists' judgements of relative pitch. *The Quarterly Journal of Experimental Psychology*, 62(11), 2156-2170. doi:10.1080/17470210902745009
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169-92. doi:15217330
- Roelofs, A., Özdemir, R., & Levelt, W. J. M. (2007). Influences of spoken word planning on speech recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 900-913. doi:10.1037/0278-7393.33.5.900
- Saltzman, E. L., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1(4), 333–382.
- Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, 23(2-3), 429-435.
- Schubö, A., Aschersleben, G., & Prinz, W. (2001). Interactions between perception and action in a reaction task with overlapping SR assignments. *Psychological Research*, 65(3), 145–157.
- Schütz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: action-induced modulation of perception. *Trends in Cognitive Sciences*, 11(8), 349-355. doi:10.1016/j.tics.2007.06.005

- Shaiman, S., & Gracco, V. L. (2002). Task-specific sensorimotor interactions in speech production. *Experimental Brain Research*, 146(4), 411-418. doi:10.1007/s00221-002-1195-5
- Smith, A. (2006). Speech motor development: Integrating muscles, movements, and linguistic units. *Journal of Communication Disorders*, 39(5), 331-349. doi:10.1016/j.jcomdis.2006.06.017
- Steeve, R. W., Moore, C. A., Green, J. R., Reilly, K. I., & McMurtrey, J. R. (2008). Babbling, chewing, and sucking: Oromandibular coordination at 9 months. *Journal of Speech, Language, and Hearing Research*, 51(6), 1390-1404. doi:10.1044/1092-4388(2008/07-0046)
- Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica*, 45(2-4), 175-197.
- Tremblay, S., Houle, G., & Ostry, D. J. (2008). Specificity of speech motor learning. *Journal of Neuroscience*, 28(10), 2426.
- Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 830-846.
- Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *The Journal of the Acoustical Society of America*, 117, 338.
- Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*, 316(5828), 1159.
- Werker, J. F. (1995). Exploring developmental changes in cross-language speech perception. In L. R. Gleitman & M. Liberman (Eds.), *Language: An invitation to cognitive science*, Vol. 1 (2nd ed.). An invitation to cognitive science (pp. 87-106). Cambridge, MA: Mit Press.
- Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, 103(1), 147-162.
- Whalen, D. H., Levitt, A. G., & Goldstein, L. M. (2007). VOT in the babbling of French- and English-learning infants. *Journal of Phonetics*, 35(3), 341-352. doi:10.1016/j.wocn.2006.10.001
- Wilson, E. M., Green, J. R., Yunusova, Y., & Moore, C. A. (2008). Task specificity in early motor development. *Seminars in Speech & Language*, 29(4), 257-266. doi:10.1055/s-0028-1103389
- Wohlschläger, A. (2000). Visual motion priming by invisible actions. *Vision Research*, 40(8), 925-930. doi:10.1016/S0042-6989(99)00239-4

- Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 397-412. doi:10.1037/0096-1523.24.2.397
- Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7-8), 1317-1329. doi:10.1016/S0893-6080(98)00066-5
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-Month-olds use distinct objects as cues to categorize speech information. *Cognition*, 113(2), 234-243. doi:10.1016/j.cognition.2009.08.010
- Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*. doi:10.1111/j.1532-7078.2009.00024.x
- Zwicker, J., Grosjean, M., & Prinz, W. (2008). A contrast effect between the concurrent production and perception of movement directions. *Visual Cognition*, 16(7), 953-978. doi:10.1080/13506280701653586
- Zwicker, J., Grosjean, M., & Prinz, W. (2010). On interference effects in concurrent perception and action. *Psychological Research*, 74(2), 152-171. doi:10.1007/s00426-009-0226-2

APPENDICES

Appendix A

This appendix contains data that supplements the empirical work outlined in Chapter 2. Here, a summary of response patterns to the acoustic /ava/ tokens are provided for both Experiments 1 and 2 of that chapter.

Experiment 1

Using the same elimination criterion as in the analysis of responses to acoustic /aba/ tokens, participants who scored less than 50% correct when identifying /ava/ in either of the baseline conditions were eliminated from the current analysis ($n = 8$). The remaining 40 participants showed similarly high proportions of correct /ava/-identification in the baseline conditions from both the Articulate ($M = .85$, $SD = .15$) and Imagine ($M = .86$, $SD = .15$) blocks. The following table displays the mean Misperception Indices (MPIs) (i.e., the percent-correct in the baseline condition minus the percent-correct score in each experimental condition) and their corresponding standard deviations for each experimental condition. As mentioned previously in Chapter 2, no effects of condition were observed in the analysis of the MPIs for acoustic /ava/.

Table 5.1 – MPis for acoustic /ava/ from Experiment 1 of Chapter 2.

Articulate Condition	/aba/	/afa/	/ava/
Mean MPI	.055	.068	.088
Standard Deviation	.169	.185	.172
Imagine Condition	/aba/	/afa/	/ava/
Mean MPI	.033	.033	.063
Standard Deviation	.202	.162	.196

Experiment 2

Again, participants who scored less than 50% correct when identifying /ava/ in either of the baseline conditions were eliminated from the current analysis ($n = 1$). The remaining 49 participants showed similarly high proportions of correct /ava/-identification in the baseline condition ($M = .94$, $SD = .10$). The following table displays the mean Misperception Indices (MPis) and their corresponding standard deviations for each experimental condition. As mentioned previously in Chapter 2, no effects of condition were observed in the analysis of the MPis for acoustic /ava/.

Table 5.2 – MPis for acoustic /ava/ from Experiment 2 of Chapter 2.

Articulate Condition	/afa/	/ aõa /	/ama/	/aya/
Mean MPI	.063	.080	.037	.067
Standard Deviation	.151	.135	.101	.157

Appendix B

A copy of the certificate of approval from the UBC Research Ethics Board for the research described in Chapters 2, 3, and 4 is included on the following page.



The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road, Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL- MINIMAL RISK RENEWAL

PRINCIPAL INVESTIGATOR: Janet F. Werker	DEPARTMENT: UBC/Arts/Psychology, Department of	UBC BREB NUMBER: H95-80023
INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:		
Institution UBC		Site Vancouver (excludes UBC Hospital)
Other locations where the research will be conducted: N/A		
CO-INVESTIGATOR(S): Laurel Fais Whitney M. Weikum Trisha R. Pranjivan Henny Yeung Krista Byers-Heinlein Stephanie Helm Lillian A. May Lawrence M. Chen Alison J. Greuel Tania Zamuner Nazanin Akmal Judit Gervain Ladan G. G. Hamadani		
SPONSORING AGENCIES: Natural Sciences and Engineering Research Council of Canada (NSERC) - "Linking speech perception to language acquisition: Biases, mechanisms, and products"		
PROJECT TITLE: Linking Speech Perception to Language Acquisition: Biases, Mechanisms and Products		

EXPIRY DATE OF THIS APPROVAL: November 26, 2010

APPROVAL DATE: November 26, 2009

The Annual Renewal for Study have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board

Dr. M. Judith Lynam, Chair
Dr. Ken Craig, Chair
Dr. Jim Rupert, Associate Chair
Dr. Laurie Ford, Associate Chair
Dr. Anita Ho, Associate Chair