The vocal-motor system of the human brain

THE VOCAL-MOTOR SYSTEM OF THE HUMAN BRAIN

BY

MICHEL BELYK

A THESIS

SUBMITTED TO THE DEPARTMENT OF PSYCHOLOGY, NEUROSCIENCE &

BEHAVIOUR

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2015) McMaster University

(Psychology, Neuroscience & Behaviour) Hamilton, Ontario, Canada

TITLE:	The vocal-motor system of the human brain
AUTHOR:	Michel Belyk B.A. (Psychology, Neuroscience & Behaviour) McMaster University, Hamilton, Ontario, Canada
SUPERVISOR:	Dr. Steven Brown
NUMBER OF PAGES:	xvi, 199

I dedicate this dissertation to Borys & Sharon Malyczewsky, who are upright citizens of the world that I can only aspire to emulate.

Abstract

The larynx is the mammalian organ of vocalization. Humans have a degree of control over this organ considerably beyond the abilities of other primates, most notably in our control over the larynx during speech. Although there is an abundance of research on the neural basis of speech, relatively little of this research has focused on the control of the larynx. First, I performed a meta-analysis to search for brain areas responsible for making explicit judgments about affective prosody to identify candidate premotor areas in prefrontal cortex that may also plan the affective component of affective prosody (Chapter 2). The inferior frontal gyrus pars orbitalis was the only prefrontal region preferentially engaged by affective vocalizations. Second, I used functional magnetic resonance imaging to determine whether there are discrete neural systems for producing innate-affective versus arbitrary non-affective vocalizations in the human brain, as has been predicted from non-human primate models (Chapter 3). The vocal-motor system demonstrated a lack of specialization since both types of vocalizations engaged the entire network. Third, I searched for brain areas that were preferentially engaged during vocal imitation (Chapter 4), which is a key process in vocal learning. Vocal imitation preferentially engaged a cortico-striate network similar to that predicted from

avian models of vocal imitation. Finally, I performed a meta-analysis to explore the neural basis of persistent developmental stuttering (Chapter 5), a speech disorder that is associated with poor control of the laryngeal muscles. Among other brain areas, primary motor regions controlling the larynx were abnormally activated in the brains of people who stutter. Together these studies advance our knowledge of the human vocal-motor system, how it relates to that in other species, and how this system may be disrupted in persistent developmental stuttering. I discuss remaining gaps in our knowledge that will be the focus of my future research.

Declaration of academic achievements

This dissertation is organized in a sandwich thesis format as approved by McMaster University. It consists of six chapters. Chapter one provides an overview of the larynx, vocal behaviour and the neural control thereof. Chapters two, three and five are articles published in peer-reviewed journals. These chapters have been reformatted to match the thesis aesthetic, but are otherwise unchanged. Chapter four is a manuscript currently under review. Chapter six provides a brief summary of the findings of the principle chapters and discusses key gaps in our knowledge of this system that outline a program for future research.

Chapter 1 – General introduction.

Author: Michel Belyk

Chapter 2 – Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies.

Authors: Michel Belyk and Steven Brown

Publication: Social Cognitive and Affective Neuroscience, 2014, 9, 1395-1403.
Comments: MB conceived of the study, collected the data, analyzed the data and wrote the first draft of the manuscript. MB and SB edited the manuscript.
Reprinted with permission.

Chapter 3 – Pitch underlies activation of the vocal system during affective vocalization.

Authors: Michel Belyk and Steven Brown

Publication: Social Cognitive and Affective Neuroscience, In Press.
Comments: MB & SB conceived of the study. MB collected the data, analyzed the data and wrote the first draft of the manuscript. MB and SB edited the manuscript.
Reprinted with permission.

Chapter 4 – The neural basis of vocal pitch imitation in humans.

Authors: Michel Belyk, Peter Q. Pfordresher, Mario Liotti and Steven Brown

Publication: Submitted to the Journal of Cognitive Neuroscience
Comments: SB, PQP and ML conceived of the study. SB collected the data. MB and SB analyzed the data. MB wrote the first draft of the manuscript. MB, PQP, ML and SB edited the manuscript.

Chapter 5 – Stuttering as a trait or state - an ALE meta-analysis of neuroimaging studies.

Authors: Michel Belyk, Shelly Jo Kraft and Steven Brown

Publication: European Journal of Neuroscience, 2015, 41, 275-284.

Comments: MB & SB conceived of the study. MB collected the data, analyzed the data and wrote the first draft of the manuscript. MB, SJK and SB edited the manuscript.

Reprinted with permission.

Chapter 6 – General discussion

Author: Michel Belyk

Acknowledgements

I first thank the members of my supervisory committee, Dr. Geoff Hall and Dr. David Feinberg for guidance and support throughout my graduate career; in particular I thank my supervisor Dr. Steven Brown for pointing me towards questions that matter and for never pulling a punch. I am grateful to my parents, all four of them, for making me who I am. To past and present members of the NeuroArts lab Dr. Ivan Chow, Pat Savage, Tom Rzeszutek, Ye Yuan, Lea Chauvigné, Blair Ellis, and Mathew Berry: thanks for sharing the adventure. My friends and colleagues Dr. Zachary Durisko, Dr. Cody Dey, Dr. Adam Reddon, Dr. Chris Teeter, Marcus Morrisey, Matt Pachai, Evan Borman, Chris Slugocki, and Chris McAllister among others sparred with me in arguments that, if not always sound, were at least loud. Friday Social made our little corner of academia a friendlier place and was the carrot urging me through every week. And not least I thank my wife, Lisa Belyk, whose patience with me is truly a source of awe.

Contents

List of Tables

List of Figures

Chapter 1

General introduction

Michel Belyk

The larynx is the mammalian organ of vocalization. It houses the vocal folds, whose vibration is the principal sound-source for vocal communication, from human universals such as laughter, crying and affective prosody to complex and culturally dependent forms of communication such as speech and song. In the latter uses of the voice, humans differ markedly from other primates. This in turn suggests differences in the vocal-motor system – the collection of brain areas that control vocal output. Although there is an abundance of research on speech, and the neural basis thereof, there is a paucity of research on the neural control of the larynx.

1.1 The anatomy of vocalization

Two different sets of muscles control laryngeal functioning: the intrinsic and extrinsic laryngeal muscles. The intrinsic laryngeal muscles modify the positioning and tension of the vocal folds internal to the larynx, whereas the extrinsic laryngeal muscles support and modify the position of the larynx within the airway (Seikel, King, & Drumwright, 2010).

The intrinsic muscles of the larynx control two dimensions of vocal-fold movement. First, the vocal folds can be adducted by contraction of the lateral cricoarytenoid, oblique interarytenoid and/or transverse interarytenoid muscles to obstruct airflow or abducted by contraction of the posterior cricoarytenoid to allow air to pass freely. In the adducted position, the passage of air causes the vocal folds to vibrate, producing the sound-source for vocal pitch. The second function of the intrinsic laryngeal muscles is to modulate vocal pitch. The tension of the vocal folds is modulated by the cricothyroid muscle (Kempster, Larson, & Kistler, 1988), which influences the frequency of vibration of the vocal folds. Contraction of this muscle lengthens and tenses the vocal folds, which raises vocal pitch (Hollien & Moore, 1960). The thyroarytenoid muscle lies within the vocal folds themselves and has a role in modulating vocal fold tension, although its relation to pitch is complex. This muscle can either lower or raise vocal pitch depending on interactions with the cricothyroid muscle (Lowell & Story, 2006; Titze, Luschei, & Hirano, 1989).

In contrast to the intrinsic laryngeal muscles, the extrinsic laryngeal muscles control the vertical position of the larynx within the airway. Two sets of

muscles pull the larynx in opposing directions along the vertical axis. Laryngeal elevators raise the larynx during swallowing and vomiting so as to protect the airway (Ardran & Kemp, 1952). These muscles extend from the larynx to more-superior structures (Seikel et al., 2010), including the mandible (i.e., mylohyoid, geniohyoid and anterior digastricus muscles), pharynx (i.e., thyropharyngeus muscle), tongue (i.e., hyoglossus and genioglossus muscles), and temporal bone (i.e., stylohyoid and posterior digastricus muscles). Laryngeal depressors, also known as the strap muscles, lower the larynx during yawning (Barbizet, 1958). These muscles extend from the larynx to more-inferior structures (i.e., Seikel et al., 2010), including the sternum (i.e., sternohyoid and sternohyroid muscles) and scapula (i.e., omohyoid muscle). However, both sets of muscles also have a secondary influence on vocal pitch by altering the relative positions of the various laryngeal cartilages, which indirectly affects the tension of the vocal folds (Vilkman, Sonninen, Hurme, & Körkkö, 1996). Indeed, vertical movement of the larynx is readily observed during pitch modulation. Untrained singers generally lower the larynx when they sing a low pitch (Roubeau, Chevrie-Muller, & Saint Guily, 1997) and raise the larynx when they sing a high pitch (Pabst & Sundberg, 1993). However, trained singers can maintain a relatively constant vertical position of the larynx across their vocal range (Shipp & Izdebski, 1975).

1.2 Vocal communication in humans

The ability to learn and modify vocal behaviour is broadly categorized into two types of abilities. Vocal usage learning is the ability to produce an existing

vocal signal in a new context as a result of experience with other individuals, while vocal production learning is the ability to modify vocal signals as a result of experience with other individuals (Janik & Slater, 2000). While vocal usage learning is common to many species, vocal production learning is relatively rare (Petkov & Jarvis, 2012).

The existence of theatre and cinema amply demonstrates that humans are capable of vocal usage learning since actors laugh, cry and intone their voices as dictated by a scene rather than their own affective states. The voice can convey a speaker's emotional state by modulating vocal acoustic parameters, such as vocal pitch, loudness, and tempo, among others (Banse & Sherer, 1996). This is referred to as affective prosody or tone of voice. Listeners can reliably recognize a broad range of vocally-expressed emotions, even when the spoken words are unrelated to the emotion being expressed (Belin et al., 2008; Fairbanks & Pronovost, 1938; Simon-Thomas et al., 2009) or when recordings are filtered to remove segmental content (Lieberman & Michaels, 1962). Unlike the words that make up the segmental aspect of speech, affective vocalizations can be recognized across languages (Laukka et al., 2013) and between cultures that have had only minimal historical contact (Sauter et al., 2010) − although with some cultural variation (Scherer & Wallbott, 1994). Indeed, infants with congenital hearing-impairments, that have therefore had little or no exposure to models of these sounds, produce affective vocalizations that are acoustically similar to those of normal-hearing infants (Scheiner et al., 2004, 2006). Together, these findings suggest that the vocal expression of emotions

is a human universal, akin to facial expressions of emotion (Ekman, Sorenson, & Friesen, 1969; Paul Ekman & Friesen, 1986), and that vocal patterns that express emotions are likely shaped by evolution, as proposed by Darwin (1872). Therefore, to produce affective vocalizations that do not match one's affective state is a demonstration of vocal usage learning.

Beyond repurposing existing signals to new contexts, humans also have the capacity for vocal production learning. Humans learn novel vocal patterns as a part of speech acquisition. Each language has a unique repertoire of phonemes that are combined to form spoken words, and voicing is a ubiquitous acoustic cue for distinguishing between these phonemes. The presence versus absence of voicing is a feature common to many languages that distinguishes between pairs of phonemes. For example, in English the phonemes /z/ and /s/ are identical in articulation but /z/ is produced with adducted vocal folds while /s/ is produced with abducted vocal folds. Speech requires a constant and irregular cycling between voiced and voiceless phonemes and hence a constant cycling between adduction and abduction of the vocal fold. Some pairs of syllables, such as /pa/ and /ba/, differ only subtly in the relative timing of voicing onset, on the order of tens of milliseconds (Lisker & Abramson, 1966).

Changes in not only the presence or absence of vocal fold vibration, but also in the frequency of vibration, are common features of speech. In tone languages, such as Cantonese and Mandarin, the pitch contour with which a word is spoken can change its meaning (Yip, 2002). In addition, all human languages use pitch modulation, among other cues, as part of sentence intonation (Ladd, 2008). For example, declarative and interrogative sentences are

distinguished by rising and falling pitch contours, respectively. A local rise in vocal pitch and loudness can focus attention on particular words within a sentence or place stress on one syllable over another, which in some cases may change the meaning of a word. For example, CONtent with the stress on the first syllable is a noun referring to that which is contained, but conTENT with the stress on the second syllable is an adjective that describes an affective state.

1.3 Vocal communication in non-human primates

Few mammalian species have the capacity for vocal production learning. Among the principal exceptions are humans, dolphins (King & Sayigh, 2013), whales (Noad, Cato, Bryden, Jenner, & Jenner, 2000), and bats (Knörnschild, Nagy, Metz, Mayer, & von Helversen, 2010). Limited evidence also suggests that elephants (Poole, Tyack, Stoeger-Horwath, & Watwood, 2005; Stoeger et al., 2012), seals (Ralls, Fiorelli, & Gish, 1985; Sanvito, Galimberti, & Miller, 2007), and mice (Arriaga & Jarvis, 2013) may be capable of vocal imitation, although evidence of vocal production learning remains sparse for these species.

While some great apes have been reported to produce novel sounds, these cases usually involve non-vocal sounds such as lip smacking, forced exhalation or whistling, rather than vocalization (Bergman, 2013; Hayes & Hayes, 1951; Wich et al., 2009), although Lameira et al. (2015) reported a possible exception. Indeed, while human infants babble frequently and imitate speech sounds

modeled by adults (Kuhl & Meltzoff, 1996), chimpanzee infants produce only chimpanzee-typical calls (Hayes & Hayes, 1951).

Experiments in which chimpanzees were cross-fostered with human foster-parents have demonstrated that even small feats of vocal learning are challenging for chimpanzees. One chimpanzee, Viki, famously learned to produce four English words after extensive operational conditioning (Kellogg, 1968). These words consisted of oral sounds that approximated human consonants combined with a single whispered vowel-like sound (Hayes, 1952; Hayes & Hayes, 1951). Notably, none of these sounds were voiced. In contrast, Washoe, who was similarly cross-fostered, learned dozens of signs in American Sign Language, used these signs reliably, and generalized the meaning of signs to new cases (Gardner & Gardner, 1985). These experiments demonstrate that, while non-human primates may learn to communicate through other modalities, their vocal repertoires are relatively fixed.

However, non-human primates do have the capacity for vocal usage learning. Operational conditioning readily elicits species-typical vocalizations in novel contexts (Koda, Oyakawa, Kato, & Masataka, 2007; Pierce, 1985). While non-human primates may be poor neuroscientific models of vocal production learning, they may nonetheless be reasonable models for vocal usage learning.

1.4 Mammalian neural models of vocal usage learning

Much of our knowledge of the neural control of the larynx comes from studies of non-human mammals and of monkeys in particular. The intrinsic laryngeal muscles − those involved in regulating voicing − are innervated by lower motor neurons contained in the nucleus ambiguus of the brainstem (Jürgens, 2002). These neurons are organized somatotopically, with the thyroarytenoid muscle located ventrally, the cricothyroid muscle located dorsally and the posterior cricoarytenoid muscle in an intermediate location that overlaps with the representations of the other muscles (Hernández-Morato et al., 2013; Yoshida, Tanaka, Saito, Shimazaki, & Hirano, 1992).

The patterning of affective vocalizations is regulated by the periaqueductal gray (PAG). This nucleus receives extensive limbic inputs, projects directly to the nucleus ambiguus and electrical stimulation of the PAG elicits species-typical affective vocalizations (Jürgens & Ploog, 1970). Lesioning the PAG abolishes affective vocal responses to environmental stimuli (Jürgens & Pratt, 1979a).

The cingulate vocalization area, located in the anterior cingulate cortex (ACC), projects to the PAG allowing cortical regulation of affective vocalizations. Stimulation of the ACC elicits species-typical affective vocalizations so long as the PAG is intact (Jürgens & Pratt, 1979b). Lesions to the ACC prevent the initiation of operantly conditioned vocalizations (Aitken, 1981; Sutton et al., 1974, 1981), but have no effect on spontaneous vocalizations in

contexts that would normally elicit these responses (Jürgens & Pratt, 1979b). Hence, the ACC in non-human primates is believed to initiate volitional, but not reflexive, species-typical affective vocalizations via projections to the PAG (Jürgens, 2002, 2009) making it a critical cortical structure for vocal usage learning and the regulation of affective vocalizations in these species.

1.5 Avian neural models of vocal production learning

Three lineages of birds, namely parrots, hummingbirds and songbirds, are capable of vocal production learning (Nottebohm, 1972). Vocal production learning has been studied most extensively in songbirds. Juvenile songbirds imitate the songs of adult tutors (Roper & Zann, 2006). Although a songbird's vocal repertoire crystalizes after the juvenile period, adult songbirds nonetheless retain some degree of sensorimotor plasticity (Tumer & Brainard, 2007). Males adapt their singing in the presence of females. A male's song performance is stereotyped when females are present, but variable when they are absent. This contextual shift has been characterized as an alternation between performance and sensorimotor practice.

The avian song system consists of two pathways: a descending vocal-motor pathway and a forebrain-striatal loop (Jarvis et al., 2005). Both pathways receive input from Area HVC (formerly known as the higher vocal center). The descending pathway consists of the robust nucleus of the arcopallidum (RA), which projects to lower motor neurons in the brainstem, located in the

tracheosyringial division of nucleus XII, which in turn project to the muscles of the vocal organ. The forebrain-striatal loop consists of three structures: Area X, the dorsolateral nucleus of the medial thalamus (DLM) and the lateral magnocellular nucleus of the nideopallium (LMAN). Inputs to this loop are received by Area X, which is a subdivision of the striatum. Area X projects to the DLM of the thalamus, which in turn projects to the LMAN. LMAN projects back to Area X to close the loop and sends output to RA, which influences the descending motor pathway.

While lesions to the descending vocal-motor pathway profoundly disrupt song production (Nottebohm, Stokes, & Leonard, 1976), lesions to the forebrain-striatal loop disrupt vocal imitation and song learning, but spare the production of songs that have already been learned (Bottjer, Miesner, & Arnold, 1984; Sohrabji, Nordeen, & Nordeen, 1990). Neurophysiological evidence suggests that neurons along the forebrain-striatal loop compute causal inverse models that map target sounds onto the motor commands that reproduce them (Giret, Kornfeld, Ganguli, & Hahnloser, 2014). However, increased variability in neural firing along the forebrain-striatal loop during undirected singing (Hessler & Doupe, 1999) results in increased song variability (Kao, Doupe, & Brainard, 2005; Liu & Nottebohm, 2005), and lesioning this pathway prevents such context-dependent changes in song variability to occur (Kao & Brainard, 2006). Ablating part of the descending pathway, such that only the forebrain-striatal loop drives vocalization, results in a reversion to the oscine equivalent of babbling, which is characterized by highly variable song (Aronov,

Andalman, & Fee, 2008). Hence, this pathway is not only critical for song acquisition in juvenile songbirds, but may also be involved in the maintenance and sensorimotor development of vocal-motor patterns in adults.

Several of the brain areas that comprise the two songbird vocal pathways have analogues in the human brain (see Jarvis, Güntürkün, & Bruce, 2005 for a review), and these analogues are also active when humans sing (Brown, Martinez, Hodges, Fox, & Parsons, 2004). A recent study sought to identify specializations in patterns of gene expression in humans and songbirds relative to their taxonomic neighbours that lack the capacity for vocal production learning (Pfenning et al., 2014). This study found shared molecular specializations in several key regions of the vocal-motor systems of these species that were strongly suggestive of anatomical and functional analogy. Nucleus RA is analogous to two areas of the human primary motor cortex that are associated with laryngeal motor control; the tracheosyringial division of nucleus XII in songbirds is analogous to the human nucleus ambiguus; Area X is analogous to the human putamen. The songbird LMAN has been speculated to be an analogue to Broca's area, although the evidence for this analogy remains sparse (Pfenning et al., 2014). The strong analogy between the human and songbird vocal-motor systems, despite the great phylogenetic distance between these species and the lack of homology with more closely related species, suggests that the brains of humans, songbirds, hummingbirds and parrots may have converged on similar solutions to vocal production learning. For this reason, these avian species may be highly useful animal models of the human vocal-motor system.

Several key species differences temper the analogy between the songbird and human vocal-motor systems. First, although vocal production learning is common to both sexes in some species of songbirds, it is sexually dimorphic in others (Riebel, 2003). Notably, zebra finches, which are the most commonly studied model system, are strongly sexually dimorphic in both singing behaviour and in the size of the nuclei in the song system (MacDougall-Shackleton & Ball, 1999). Second, in adult songbirds damage to Area X has only a limited effect on the production of songs that are already in a bird's vocal repertoire (Aronov et al., 2008; Bottjer et al., 1984; Sohrabji et al., 1990), while in humans degenerative diseases of the basal ganglia, such as Parkinson's disease, can cause severe hypophonia Blumin, Pcolinsky, & Atkins, 2004; Canter, 1963). Finally, the mammalian larynx and avian syrinx differ markedly in structure. While the larynx is a single midline structure that sits above the trachea, the syrinx is a paired structure found in the bronchi that may produce sound independently of one another. Indeed, while nucleus RA projects predominantly to ipsilateral brainstem nuclei, the number of projections to the contralateral side varies strongly between species and may be related to hemispheric dominance for singing (Wild, Williams, & Suthers, 2000).

1.6 The human vocal-motor system

On the basis of the animal models reviewed above, the human vocal-motor system may consist of three neural pathways that together control vocalization. First, the ACC-PAG pathway, that is critical for affective vocalization and vocal usage learning in non-human primates, appears to be conserved in

humans (Barrett, Pike, & Paus, 2004; Wattendorf et al., 2013). However, the specificity of this pathway for affective vocalization is uncertain, since brain imaging experiments demonstrate that this pathway is activated during vocalizations that express no affect (e.g., Aziz-Zadeh, Sheng, & Gheytanchi, 2010; Schulz, Varga, Jeffires, Ludlow, & Braun, 2005). Second, the forebrain-striatal loop, which is critical for vocal production learning in songbirds, has been hypothesized to have analogues in the human brain (Jarvis, 2007). However, no previous research has examined this putative functional analogy between vocally imitating birds and humans despite common specializations in gene expression profiles (Pfenning et al., 2014) that provide some evidence for evolutionary convergence between these species. Finally, the primary motor cortex contains a somatotopic representation of the larynx that is more developed in humans than in other primate species. This section reviews the literature surrounding the latter pathway in detail.

The larynx-controlling region of the precentral gyrus forms an evolutionary continuum from monkeys through non-human primates to humans. In monkeys, the cortical larynx area is located in the premotor cortex, rather than primary motor cortex. Electrical stimulation of the monkey cortical larynx area stimulates contraction of the intrinsic and extrinsic laryngeal muscles (Hast et al., 1974), but does not elicit vocalization (Jürgens, 1974). Although neurons in this region are active during vocalization (Coudé et al., 2011), bilateral lesions have little effect on vocal behaviour (Jürgens et al., 1982; Kirzinger & Jürgens, 1982) and this region projects only indirectly to lower motor neurons

in nucleus ambiguus via relays on the reticular formation (Jürgens & Ehrenreich, 2007). In great apes, the cortical larynx area has migrated towards the primary motor cortex, and electrical stimulation elicits sound production. Although lesions to this area do not affect vocal behaviour in non-human apes (Kuypers, 1958b), a sparse population of neurons do project directly from the larynx controlling region to the nucleus ambiguus (Kuypers, 1958b).

In humans, two pericentral cortical areas, at the dorsal and ventral extremes of the orofacial somatotopic division of primary motor cortex, control the larynx and vocalization (Bouchard, Mesgarani, Johnson, & Chang, 2013). The more dorsal of these areas is sometimes referred to as the larynx phonation area (LPA; Brown, Ngan, & Liotti, 2008) due to its involvement in both phonation and non-phonatory laryngeal movements. The more ventral of the larynx motor areas is located either superficially on the subcentral gyrus (Bouchard et al., 2013) or within the adjacent Rolandic operculum (Brown et al., 2008) slightly posterior to the location of the larynx area in non-human primates (Hast et al., 1974; Leyton & Sherrington, 1917). Electrical stimulation of either region elicits sound production in humans (Foerster, 1931; Penfield & Boldrey, 1937). Lesions to the orofacial primary-motor region can result in muteness (Jürgens et al., 1982), and direct projections from this region to the nucleus ambiguus are more abundant in humans than in other species (Iwatsubo, Kuzuhara, & Kanemitsu, 1990; Kuypers, 1958a).

Early electrophysiological studies in humans revealed that stimulation of the dorsal larynx area, unlike stimulation of larynx motor cortex in monkeys (Hast et al., 1974), elicits sustained vocalization resembling a vowel sound

(Breshears, Molinaro, & Chang, 2015; Penfield & Boldrey, 1937). With the advent of modern brain imaging technologies, studies using functional Magnetic Resonance Imaging (fMRI) and Position Emission Tomography (PET) have since characterized the dorsal larynx area as a cortical larynx area for the control of the intrinsic musculature of the larynx. This region is active not only during simple vocalization (Brown et al., 2004), but also during forced expiration (Loucks, Poletto, Simonyan, Reynolds, & Ludlow, 2007; Simonyan, Saad, Loucks, Poletto, & Ludlow, 2007) and vocal fold adduction (Brown et al., 2008), both of which are motoric components of vocalization. Importantly, brain imaging research has established that the dorsal larynx area is distinct from the adjacent somatotopic motor representations of the articulatory muscles, namely the lips and tongue (Brown et al., 2008; Grabski et al., 2012; Takai, Brown, & Liotti, 2010) although it is also active in conjunction with these neighbouring regions in speech (Turkeltaub, Eden, Jones, & Zeffiro, 2002) and song (Brown et al., 2009). Expertise in singing is associated with increased activation in primary somatosensory cortex immediately posterior to the dorsal larynx area (Kleber, Veit, Birbaumer, Gruzelier, & Lotze, 2010), and the level of song-related activation in this area is modulated by applying a local anesthetic to the vocal folds (Kleber, Zeitouni, Friberg, & Zatorre, 2013). Transcranial magnetic stimulation experiments suggest that the dorsal larynx area may contain spatially distinct representations of the cricoarytenoid and thyroarytenoid muscles (Rödel et al., 2004), which raise and lower vocal pitch, respectively. However, the spatial trend observed by Rödel et al. (2004) is a complete reversal of the somatotopic organization in other primate species

(Hast et al., 1974), suggesting a need for replication.

The human motor homunculus contains a second laryngeal representation in the subcentral gyrus, or adjacent Rolandic operculum, near the location of larynx-motor cortex in non-human primates. Unlike the dorsal larynx area, stimulation of the ventral larynx area does not elicit sustained vocalization. However, stimulation in this region does elicit grunting sounds (Penfield & Boldrey, 1937). Foerster (1931) described these sounds as co-occurring with rhythmic movements of the lips, tongue, mandible and pharynx. This suggests a possible role of the subcentral gyrus in swallowing. Indeed activations have been observed in a similar location during swallowing (Martin & Goodyear, 2001; Martin et al., 2004), although no study has formally tested the association between vocalization and swallowing-related cortical motor representations in the subcentral gyrus. Brain imaging studies have observed activation of this region, along with the dorsal larynx area, during vocalization, glottal stops (Brown et al., 2008) and forced exhalation (Loucks et al., 2007). A recent study recorded neural activity in the precentral gyrus with an implanted high-density multi-electrode array in waking patients while they produced monosyllabic vocalizations (Bouchard et al., 2013). The phonatory component of syllable production was associated with increased local field potentials in both the dorsal and ventral larynx areas. Hence, in addition to the migration of the primate larynx motor area towards primary motor cortex, humans may have evolved a second motor representation of the larynx de novo.

1.7 Stuttering as a model of disordered laryngeal control

One speech disorder that may highlight the consequences of disordered control of the larynx is persistent developmental stuttering. This disorder is characterized by speech with involuntary repetitions, prolongations, hesitations and blocks at the levels of syllables and words (Wingate, 1964). This disorder is associated with poor coordination of the musculature responsible or adducting and abducting the vocal folds (Freeman & Ushijima, 1978), stuttering is more likely to occur at the onset of voicing (Adams & Reis, 1971), and many alterations to speech patterns that ameliorate stuttering reduce the frequency of vocal fold adduction/abduction (Ingham, Bothe, Wang, Purkhiser, & New, 2012). An early meta-anlysis of brain imaging studies revealed that persistent developmental stuttering is associated with abnormal activation of the dorsal larynx motor area (Brown, Ingham, Ingham, Laird, & Fox, 2005). Although the neural correlates of stuttering vary greatly between individuals (Wymbs, Ingham, Ingham, Paolini, & Grafton, 2013), those cases with laryngeal involvement may act as a model of vocal-motor disruption.

1.8 Research objectives

With the advent of modern brain imaging technologies, such as fMRI and PET, the last three decades have seen a proliferation of research on the neural basis of speech. However, relatively little research has been conducted on the laryngeal component of speech, despite its key role as the principal sound

source for vocalization. The objectives of this dissertation are threefold: i) to localize potential premotor areas involved in the affective component of speech, ii) to examine the activity of the vocal-motor system during tasks of vocal usage learning and vocal production learning in light of predictions from animal models, and iii) to establish persistent developmental stuttering as a potential model of vocal-motor malfunction.

Chapter 2

Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies

Michel Belyk & Steven Brown.

Social Cognitive and Affective Neuroscience, 2014, 9, 1395-1403.

2.1 Abstract

Prosody refers to the melodic and rhythmic aspects of speech. Two forms of prosody are typically distinguished: "affective prosody" refers to the expression of emotion in speech, whereas "linguistic prosody" relates to the intonation of sentences, including the specification of focus within sentences and

stress within polysyllabic words. While these two processes are united by their use of vocal pitch modulation, they are functionally distinct. In order to examine the localization and lateralization of speech prosody in the brain, we performed two voxel-based meta-analyses of neuroimaging studies of the perception of affective and linguistic prosody. There was substantial sharing of brain activations between analyses, particularly in right-hemisphere auditory areas. However, a major point of divergence was observed in the inferior frontal gyrus: affective prosody was more likely to activate Brodmann area 47, while linguistic prosody was more likely to activate the ventral part of area 44.

Keywords: affective prosody, linguistic prosody, speech, emotion, ALE meta-analysis, brain imaging, inferior frontal gyrus

2.2 Introduction

Prosody comes from the Greek prosodia, meaning "sung to music" (Pearsall, Hanks, Soanes & Stevenson, 2005). Speech prosody therefore refers to the song-like vocal modulations that accompany speech. For this reason, it is often considered to be "the music of speech" (Wennerstrom, 2001). The pitch modulations associated with speech prosody convey two broad categories of information. On the one hand, pitch modulations convey information about a speaker's emotional state (Fairbanks & Pronovost, 1938), what has been referred to as "emotional" or "affective" prosody (Monrad-Krohn, 1947). On the other hand, they provide cues regarding syntax and pragmatics (Beach, 1991),

what has been referred to as "intrinsic" or "linguistic" prosody (Monrad-Krohn, 1947). While these two types of prosody are functionally distinct, they rely on a common set of acoustic cues related to pitch, loudness, tempo, and voice quality (Fonagy, 1978; Juslin & Laukka, 2003). The sharing of acoustic parameters by these two processes suggests that they might rely on a common system for the perception of pitch but that this pitch information may be fed into distinct systems for processing either emotion (affective prosody) or syntax/pragmatics (linguistic prosody).

Affective prosody conveys a speaker's emotional state largely through global changes in pitch height and loudness, although other acoustic features also serve to disambiguate emotional states (Banse and Sherer, 1996). Emotional expressions can take the form of "affect bursts" (Schröder, 2003) that have emotional but not semantic meaning (e.g., "Yuck!") or can occur concurrently with normal speech. Affective prosody conveys a broad range of emotional states (Sauter and Scott, 2007) that can be recognized across cultures without prior experience (Sauter et al., 2010; Scherer et al., 2001), much like facial expressions (Ekman et al.,1969).

Linguistic prosody uses local increases in pitch height and/or loudness to signal features like word stress (e.g., CONtent vs. conTENT; Gay, 1978), sentence focus (e.g., two WHITE shirts vs. TWO white shirts; Ladd and Morton, 1997), segmentation of the speech stream into phrases (Jusczyk et al., 1992), broad pragmatic categories of utterances (modality), such as declarative vs. interrogative sentences (Xu and Xu, 2005), and the standard intonational melodies that are used as part of mother-infant communication

(Fernald, 1992) as well as communication between adults. The conventions of linguistic prosody vary across languages and are important contributors to the melody and rhythm of speech. It is for this reason that deviations from standard prosody contribute to the impression of a foreign accent (de Mareüil and Vieru-Dimulescu, 2006).

Early investigations into the neural basis of speech prosody analyzed neurological cases of patients suffering from strokes. These studies focused overwhelmingly on the lateralization of prosody, especially compared to the well-accepted left-hemisphere dominance for the lexicosyntactic aspect of language. A major finding of these early studies was that the perception of affective prosody was impaired in patients with unilateral right-hemisphere lesions (Gorelick and Ross, 1987; Ross, 1981). However, these studies did not examine patients with left-hemisphere lesions, and studies that have since done so have reported deficits in patients with both types of unilateral lesions (Pell, 1998; Trauner et al., 1996). Similarly, deficits in the perception of linguistic prosody have been reported in patients with lesions in both the left (Pell and Baum, 1997) and right (Weintraub et al., 1981) hemispheres. A meta-analysis of this literature revealed that both affective and linguistic prosody are impaired by damage to either hemisphere, although damage to the right hemisphere tends to have a larger impact on affective prosody and the left hemisphere on linguistic prosody (Witteman, Van Ijzendoorn, Van de Velde, Van Heuven, & Schiller, 2011).

Neurological studies have generally been conducted with patients having a diverse set of lesions and have seldom reported the location of lesions beyond

the level of the hemisphere or lobe. Therefore, the neurological literature does not permit an examination of localization hypotheses at a finer scale than the lobe. Interestingly, transcranial magnetic stimulation of healthy individuals can induce deficits in the perception of affective prosody when applied to either the left or right inferior frontal gyrus (IFG; Hoekert et al., 2010).

Neuroimaging studies have been similarly inconclusive with respect to the hemispheric lateralization of prosody perception. The literature has variably reported unilateral or bilateral activations for affective prosody (Bach et al., 2008; Ethofer et al., 2009; Wildgruber et al., 2005) and linguistic prosody (Meyer et al., 2002; Strelnikov et al., 2006). Despite these inconsistencies in lateralization, neuroimaging studies have contributed to the broader localizationist account of prosody perception.

Neural models of affective-prosody perception (Ethofer et al., 2006; Schirmer and Kotz, 2006) suggest that low-level acoustic analyses are performed in the posterior superior temporal gyrus (STG) − in what has been called the "emotional voice area" (Ethofer et al., 2012) − and the superior temporal sulcus (STS). Similarly, more recent models suggest that acoustic processing is performed in the middle part of the superior temporal sulcus (mSTS; Belin et al., 2000), that identification of vocally expressed emotions is performed in either the anterior (Kotz & Paulmann, 2011) or posterior (Brück, Kreifelts, & Wildgruber, 2011) STG/STS, and that explicit evaluation of vocally-expressed emotions is performed by inferior frontal regions (Wildgruber et al., 2009). Passive perception of prosody reliably activates the STG (Dietrich et al., 2008; Humphries et al., 2005). Posterior temporal areas are proposed to project

to inferior frontal regions for explicit evaluation of emotional meaning when such evalution is task-relevant. While studies of both affective and linguistic prosody routinely report activations in Broca's area (Gandour et al., 2003a; Gandour et al., 2003b), Schirmer and Kotz (2006) proposed that a region anteroventral to Broca's area − the IFG pars orbitalis (Brodmann area [BA] 47) − may be specifically involved in the perception of affective prosody. A meta-analysis of the imaging literature on the perception of affective prosody supports the involvement of the IFG pars orbitalis when attention is directed towards affective prosody rather than away from it and the IFG pars triangularis (BA 45) whether or not attention is directed towards affective prosody (Witteman, Van Heuven, and Schiller, 2012).

The perception of prosody stimulates additional regions beyond the superior temporal and inferior frontal gyri (Brück et al., 2011; Buchanan et al., 2000). Studies of affective and linguistic prosody routinely report activations in speech-related areas − even when contrasted with other speech-perception tasks − including the anterior cingulate cortex (ACC; Doherty et al., 2004; Frühholz et al., 2011), inferior parietal lobule (IPL; Gandour et al., 2003a; Johnstone et al., 2006), anterior insula (Ethofer et al., 2009; Meyer et al., 2002), and basal ganglia (Bach et al., 2008; Meyer et al., 2004).

Given the inconsistencies in both the neurological and neuroimaging literatures, we sought to clarify the localization of prosody perception in the brain by performing a statistical meta-analysis of published neuroimaging studies of affective and linguistic prosody either separately, in contrast, or in conjunction using the "activation likelihood estimation" (ALE) method (Eickhoff

et al., 2011; Turkeltaub et al., 2002). The goal was to assess whether these two functions are mediated by shared or distinct brain networks. The major predictions were that these functions should show commonalities in posterior temporal areas that process the acoustic features of vocal pitch, but that differences should be seen in higher-level areas in the frontal lobe that generate distinct interpretations of these pitch modulations.

2.3 Methods

2.3.1 Inclusion criteria

A meta-analysis of published neuroimaging studies of affective and linguistic prosody was performed using ALE meta-analysis (Turkeltaub et al., 2002) in order to compare areas of brain activation across these functions. Published articles were retrieved in February 2012 by searches in the Web of Knowledge database using the search terms "prosody + fMRI" and "prosody + PET". The reference sections of resultant studies were searched for additional studies. Experiments in which subjects made emotional judgments were classified as "affective prosody", while studies in which subjects made judgments based on word stress, focus, syntax, or modality were classified as "linguistic prosody".

Our inclusion criteria for the studies were: 1) that brain scanning was performed using either functional magnetic resonance imaging (fMRI) or positron emission tomography (PET); 2) that papers reported activation foci in the form of standardized stereotaxic coordinates in either Talairach or Montreal Neurological Institute (MNI) space; 3) that subjects were healthy adults (thereby

excluding results from clinical populations); 4) that subjects made active judgments about the affective or linguistic prosody of auditorily-presented speech stimuli; 5) that the analyses included a high-level contrast against a suitable control condition so as to remove the influence of low-level phonological processing (e.g., passive listening or gender discrimination); and 6) that results from the entire scanned volume were reported (thereby excluding studies reporting region-of-interest analyses only). Due to the large number of studies with only partial brain coverage, we performed a separate analysis with the additional criterion 7) that the entire brain-volume was imaged (thereby excluding studies with an insufficient field of view to encompass the whole brain). This criterion is discussed further in section 2.2.

Our searches yielded 29 independent experiments conducted in German, English, French, Mandarin, Japanese, and Russian (see Supplementary Tables 1 and 2 for details). Wherever studies reported multiple experiments from the same group of subjects, the contrasts were included together as a single study. Similarly, for studies that reported the results of more than one subject-group, each group was treated separately, in accordance with the approach of Turkeltaub et al. (2011). Separate analyses were conducted for affective prosody (n=19 experiments) and linguistic prosody (n=10). GingerALE 2.1 was used for all analyses and to convert MNI coordinates to Talairach coordinates. The ALE results were registered onto a Talairach-normalized template brain using Mango (ric.uthscsa.edu/mango). All analyses were corrected for multiple comparisons using the False Discovery Rate $p < 0.05$ and cluster threshold $k = 10$.

2.3.2 Brain coverage

ALE meta-analysis is an empirical technique for the analysis of brain imaging studies (Turkeltaub et al., 2002). Each focus of activation is modeled as a three-dimensional Gaussian probability distribution whose width is determined by the size of the subject-group so as to reflect increasing uncertainty with decreasing sample size (Eickhoff et al., 2009). Maps of activation likelihoods are created for each study by taking the maximum probability of activation at each voxel. A random-effects analysis tests for the convergence of activations across studies against a null hypothesis of spatially independent brain activations.

Due to the limited brain coverage of many of the studies included in our dataset, we modified the standard ALE method in order to test the null hypothesis of spatially independent brain activations within the brain volume that was imaged in all of the included studies. Standard ALE analyses mask the brain volume to grey matter. Activation foci are unlikely to originate from ventricles or white matter. Therefore, in order to avoid skewing the empirical null distribution − and overestimating any effects − this portion of brain space must be excluded (Eickhoff et al., 2009). Similarly, activation foci cannot originate from outside the field of view for a given study, and so this region must therefore be excluded from the analysis. We therefore further restricted the analyses to the portion of the brain-volume that was imaged in all studies meeting our inclusion criteria. This area extended from z = -6 to z = 38 in Talairach space (see horizontal red lines in Figure 1). A second set of whole-brain analyses was performed to assess convergence beyond the restricted volume of coverage. Whole-brain analyses only included studies that

imaged the entire brain volume. This additional inclusion criterion reduced the number of studies to 10 and 4, respectively, for affective and linguistic prosody. For all figures and tables, all 29 experiments contributed to analyses within the restricted range. Only those experiments with full brain coverage contributed to analyses outside this range.

2.3.3 Conjunction of contrasts

In addition to separate analyses, we performed a statistical conjunction (Nichols et al., 2005) of the meta-analyses in order to determine which areas, if any, were common to affective and linguistic prosody. Direct contrasts were performed to determine which areas were specific to each of these two functions. Because there were many more studies of affective prosody than linguistic prosody in the dataset − which may bias the results − we also report the number and percentage of studies of affective prosody and linguistic prosody that contribute to each of the ALE foci. Due to the small number of studies covering the whole brain, direct contrasts are reported for the restricted analysis only.

2.3.4 Post hoc analysis of working memory demands based on task-type

The studies included in the meta-analyses used tasks that fall into two broad classes: identification tasks and same/different tasks. Subjects performing an identification task are presented with an auditory stimulus and are required to identify − from a limited set of possible responses − which emotion or

intonation is being presented. Subjects performing a same/different task are presented with pairs of stimuli and are required to indicate whether the same emotion or intonation occurs in both presentations. To the extent that the latter task requires subjects to maintain a representation of the first stimulus-presentation long enough to perform a comparison with the second, it may impose greater demands on working memory than an identification task. Among the studies included in our meta-analyses, affective prosody experiments were much more likely to use identification tasks or similar tasks with a low working memory load (16 out of 19), while studies of linguistic prosody were more evenly divided (4 and 6 low and high working memory load, respectively). We therefore compared experiments of linguistic prosody containing putatively low vs. high working memory load as, estimated from task demands, in order to account for areas of convergence that may be more reflective of working memory demands than prosody perception per se.

2.4 Results

We performed individual ALE analyses of affective and linguistic prosody. Due to the preponderance of studies with functional coverage limited to the perisylvian region alone, we performed two parallel analyses for each function, one restricted to the volume covered by all studies in the dataset (in order to avoid violating the assumptions of the ALE method) and a second, whole-brain analysis exclusively for those studies that reported whole-brain coverage. Results from both the restricted and whole brain analyses are combined in all figures and tables. Figure 1 presents the location of the major ALE foci for

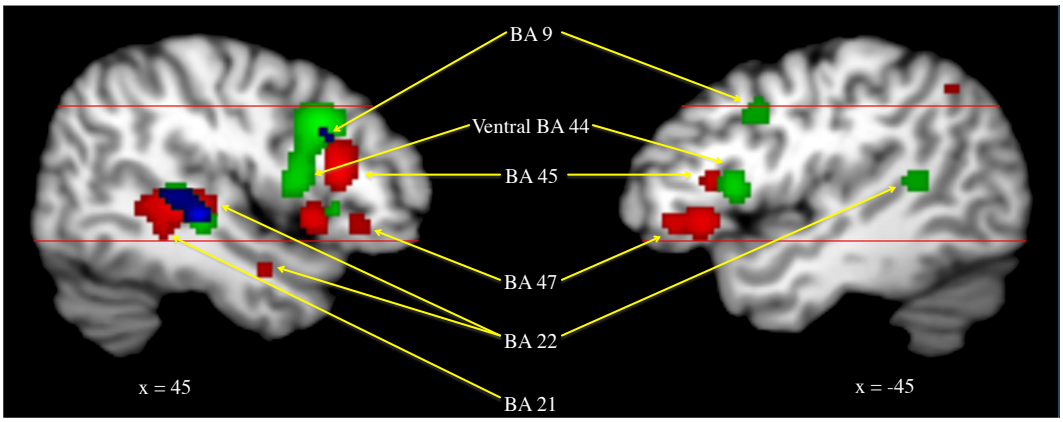

Figure 2.1: Overlap of affective and linguistic prosody.

Sagittal sections showing major foci for the individual ALE meta-analyses for affective prosody (red) and linguistic prosody (green) as well as the statistical conjunction of the two (blue). These slices demonstrate the bilateral involvement of inferior frontal regions for affective and linguistic prosody perception. The figure also demonstrates the clear segregation of functions within the inferior frontal gyrus as well as sharing in the right auditory cortex. Red lines demarcate the limits of the "restricted" analysis (z = -6 to z = 38): foci within the red lines were generated by the restricted analysis (which included all studies), while foci outside the red lines were generated by the whole-brain analysis (including only those studies that reported whole-brain coverage).

each analysis, and Table 1 provides the Talairach coordinates and cluster sizes for each ALE focus. Results will first be presented for analyses of each function separately, followed by a conjunction of analyses to identify shared regions, and finally direct contrasts to identify regions specific to each function.

Affective prosody activated both audio-vocal and limbic areas. Audio-vocal activations were observed in right pSTG, bilateral aSTG, supplementary motor area (SMA), IFG pars opercularis (BA 44), pars triangularis (BA 45) and supramarginal gyrus, right middle temporal gyrus (MTG), cerebellum, and middle frontal gyrus (BA 9, BA 10), left caudate nucleus and thalamus. Presumed emotion-related activations were observed in limbic areas, including

bilateral IFG pars orbitalis (BA 47), left amygdala, ventral anterior insula and ventral putamen, right parahippocampal gyrus (BA 28), and subcallosal gyrus (BA 34). Importantly, the frontal language areas that are most widely discussed in this literature showed bilateral activity.

Table 2.1: Affective and linguistic prosody. Talairach coordinates of likelihood maxima and cluster sizes for individual ALE analyses of affective prosody and linguistic prosody perception, respectively. Results from both the restricted and whole-brain analysis are combined. IFG: inferior frontal gyrus; MOG: middle occipital gyrus; MTG: middle temporal gyrus; PHG: parahippocampal gyrus; SMA: supplementary motor area; SMG: supramarginal gyrus; STG: superior temporal gyrus.

	Affective Prosody				Linguistic Prosody			
Brain Region	x	y	z	ALE (x10^3)	x	y	z	ALE (x10^3)
Right Hemisphere								
Frontal Lobe								
IFG pars triangularis (BA 45)	46	22	16	21.04				
Insula (BA 13)					54	-36	20	19.41
					42	8	12	13.76
IFG pars orbitalis (BA 47)	48	14	0	18.49	46	20	2	8.61
	38	26	0	15.87				
Middle frontal gyrus (BA 9)	48	16	28	11.34	46	14	30	19.94
SMA (BA 6)	8	18	50	9.04	4	16	48	14.69
					8	26	42	6.98
Middle frontal gyrus (BA 10)	34	36	10	15.34				
IFG pars opercularis (BA 44)	54	8	6	11.49	48	8	15	1.30
Temporal Lobe								
aSTG (BA 22)	54	0	4	13.77				
pSTG (BA 22)	48	-24	4	20.68	46	-24	0	12.25
	46	-32	4	20.60				
	56	-44	4	11.98				
Heschl's gyrus (BA 41)					48	-32	8	11.71
PHG (BA 28)	16	-10	-12	11.41				
MTG (BA 21)	44	-4	-16	8.95				
Subcallosal gyrus (BA 34)	26	6	-10	8.08				
Parietal Lobe								
SMG (BA 40/7)	36	-54	46	10.91	36	-58	48	15.74
Subcortical								
Claustrum					26	16	4	13.57
Cerebellum	18	-64	-16	9.42	2	-70	-10	6.75
Left Hemisphere								
Frontal Lobe								
IFG pars orbitalis (BA 47)	-40	22	-2	19.34				
	-44	34	-2	14.64				
	-50	20	0	13.57				

Continued on next page

	Affective Prosody				Linguistic Prosody			
Brain Region	x	y	z	ALE (x10^3)	x	y	z	ALE (x10^3)
IFG pars triangularis (BA 45)	-46	22	12	13.62				
Anterior insula (BA 13)	-32	22	2	16.75	-32	18	6	9.48
Middle frontal gyrus (BA 9)					-40	6	34	16.84
IFG pars opercularis (BA 44)	-42	2	6	12.70	-44	14	10	13.45
Temporal Lobe								
aSTG (BA 22)	-50	10	2	13.32				
pSTG (BA 22)					-48	-46	12	11.73
Parietal Lobe								
SMG (BA 40)	-30	-50	38	10.87	-30	-50	40	12.35
					-36	-44	38	11.86
SMG (BA 40/7)	-46	-56	42	7.70				
Subcortical								
Amygdala	-18	-6	-12	21.72				
Caudate nucleus	-12	-4	14	13.04	-16	16	8	11.23
Putamen	-22	14	-12	8.22				
Cerebellum					-28	-60	-21	6.45
Cerebellum					-6	-74	-18	6.75
Thalamus	-8	-6	10	12.60				
Midline								
Frontal Lobe								
SMA (BA 6)	0	14	48	11.99				
Occipital Lobe								
Cuneus (BA 17)					0	-82	8	16.58

In contrast to this limbic profile for affective prosody, linguistic prosody showed ALE foci primarily in speech and language areas. These included bilateral IFG pars opercularis (BA 44), pSTG, supramarginal gyrus (BA 40), middle frontal gyrus, right SMA, IFG pars orbitalis, primary auditory cortex (BA 41) and the left caudate nucleus. Non-language-related foci were observed in the bilateral insula and cerebellum as well as in the right claustrum and primary visual cortex. As with affective prosody, the ALE foci in frontal perisylvian language areas were present bilaterally.

Next, we compared the functions using conjunctions so as to identify areas of overlap versus areas of function-specificity (see Figure 1 and Table 2). Conjunction analyses demonstrated that affective prosody shared common areas

Table 2.2: Conjunction of affective and linguistic prosody.

Statistical conjunction demonstrates areas of commonality between affective prosody and linguistic prosody.

	LP ∩ AP			
Brain Regions	x	y	z	Size (mm^3)
Right Hemisphere				
Superior Temporal Gyrus (BA 22)	46	-24	0	799
Supramrginal gyrus (BA 40/7)	36	-54	46	437
Middle frontal gyrus (BA 9)	48	16	30	115
Left Hemisphere				
Supramrginal gyrus (BA 40)	-30	-50	38	27
Anterior insula (BA 13)	-30	20	4	81
Midline				
Supplementary motor area (BA 6)	0	16	48	669

with linguistic prosody. As predicted, affective and linguistic prosody showed overlapping activations in the right superior temporal gyrus (BA 22). Other areas of overlap included the bilateral supramarginal gyrus, right middle frontal gyrus, left insula, and midline SMA.

In order to identify regions that were specific to each condition, we performed direct contrasts (see the right panel of Figure 2 and Table 3). Affective prosody had a stronger association with activation in the left IFG pars orbitalis (BA 47) and thalamus as well as right pSTG (BA 22) and MTG (BA 21). Linguistic prosody had a stronger association with activation in the left pSTG, bilateral middle frontal gyrus (BA 9), bilateral IFG pars opercularis (BA 44), right supramarginal gyrus (BA 40), claustrum, and midline visual cortex.

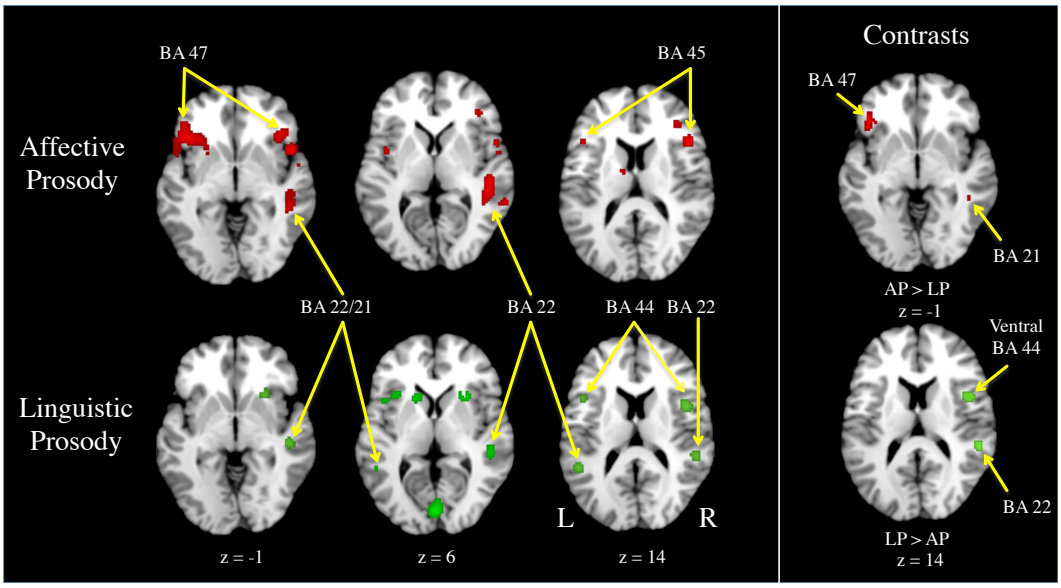

Figure 2.2: Affective vs linguistic prosody

The left panel shows the ALE foci for affective prosody (AP, red) and linguistic prosody (LP, green) registered onto axial sections. The right panel shows two direct contrasts, and highlights areas unique to each function. Affective prosody is uniquely associated with the IFG pars orbitalis (BA 47), while linguistic prosody is uniquely associated with the ventral IFG pars opercularis, (BA 44).

Table 2.3: Affective versus linguistic prosody.

Pairwise contrasts demonstrate areas of activation unique to affective prosody and linguistic prosody. *Peaks from either condition may be differentially localized as suggested by the left panel of figure 1. **May be mislocalized from nearby Putamen and/or Insula.

AP > LP	x	y	z	Size (mm^3)	AP Studies	LP Studies
Right Hemisphere						
Superior temporal sulcus (BA 22)*	44	-40	0	75	9 (47%)	5 (50%)
Middle temporal gyrus (BA 21)	44	-44	2	27	5 (26%)	1 (10%)
Left Hemisphere						
IFG pars orbitalis (BA 47)	-40	28	-6	1063	9 (47%)	1 (10%)
Thalamus	-12	-6	16	197	3 (16%)	0 (0%)
LP > AP	x	y	z	Size (mm^3)	AP Studies	LP Studies
Right Hemisphere						
Middle frontal gyrus (BA9)	50	8	30	2773	4 (21%)	5 (50%)
Angular gyrus (BA 40)	50	-32	20	1437	4 (21%)	5 (50%)
Claustrum**	28	20	6	397	0 (0%)	0 (0%)
IFG pars opercularis (BA 44)	46	8	16	111	3 (16%)	9 (90%)
Left Hemisphere						
Middle frontal Gyrus (BA 9)	-42	8	30	321	4 (21%)	3 (30%)
IFG pars opercularis (BA 44)	-40	14	8	129	4 (21%)	4 (40%)
Superior temporal gyrus (22)	-50	-42	10	45	4 (21%)	4 (40%)
Midline						
Cuneus (BA 17)	0	-76	10	1349	4 (21%)	3 (30%)

Table 2.4: Meta-analysis of verbal working memory.

Studies of linguistic prosody were divided into those with high vs low working-memory demands based on task type. The bilateral middle frontal gyrus and right STG are more likely to be reported in studies with high verbal working-memory demands. Contrasts between individual ALE meta-analyses must be interpreted cautiously to avoid falsely attributing foci in these areas to prosody perception. The number of studies with low and high working memory loads contributing to each locus corroborates the ALE results.

				High> Low Working Memory		
Brain Regions	x	y	z	ALE (10^3)	Same/Different	Identification
Right Hemisphere						
Superior Temporal Gyrus (BA 22)	48	-30	12	7.80	5 (83%)	0 (0%)
Middle frontal gyrus (BA 9)	52	12	32	6.98	4 (67%)	1 (25%)
Left Hemisphere						
Middle frontal gyrus (BA 9)	-42	14	32	10.96	2 (33%)	0 (0%)

As an additional analysis, we divided the studies of linguistic-prosody perception into those with putatively high versus low verbal working-memory load, as estimated by task demands (Table 4). Higher working memory load was associated with increased activation in the bilateral middle frontal gyrus (part of the dorsolateral prefrontal cortex) and right STG. Given that the literatures under review here were not orthogonal with respect to working memory demands, differences between individual ALE analyses in the right STG and middle frontal gyri should be interpreted with caution.

2.5 Discussion

The primary objective of this study was to use meta-analytic techniques to help clarify which brain regions are reported consistently in studies of affective and linguistic prosody perception in light of inconsistency and confusion in both

the neurological and neuroimaging literatures. We examined the functional neuroimaging literatures related to affective and linguistic prosody individually and then jointly using conjunction and contrast methods. The results revealed both shared and distinct components of the networks involved in these processes, reflecting both the perception of vocal-pitch modulation and its functional interpretation.

Our strongest prediction of overlap between the two functions was for auditory association areas in the pSTG. Interestingly, the right pSTG has been dubbed the "emotional voice area" by researchers of affective prosody (Ethofer et al., 2012). In confirmation of this area's role in emotional voice perception, we observed convergence centered in right Heschl's gyrus and extending into the pSTG for studies of affective prosody. However, we observed a similar area of convergence bilaterally for linguistic prosody as well as for verbal working memory, and the pSTG is commonly reported in studies of music perception as well (Brown et al., 2004; Zatorre et al., 1994). Indeed, Wiethoff et al. (2008) observed that activation in this region could be explained entirely by the acoustic parameters of the stimuli. The pSTG appears to respond to a variety of types of auditory stimuli and may not be specific to emotional voices. Emotional voices may simply contain a larger degree of pitch modulation than the neutral voices that are typically used as baseline stimuli in many studies of affective prosody.

In addition to demonstrating overlap in right auditory areas, the conjunction analysis revealed convergence across functions in the SMA, a motor structure involved in speech production. Electrical stimulation of the anterior portion of the left (but not right) SMA elicits vocalization (Fried et al., 1991). Lesions to this area can cause aphasic symptoms (Fontaine et al., 2002) and akinetic mutism (Nagaratnam et al., 2004). The individual ALE meta-analyses of affective prosody and linguistic prosody suggested a role of the right SMA in those functions as well. Surprisingly, convergence between studies was not observed in the ACC just ventral to the SMA. Animal models of vocal control demonstrate an important role of the ACC in top-down control of the vocal-motor nuclei in the brain stem (Jürgens, 2002), and ALE meta-analysis of human neuroimaging studies of vocalization show foci in this region for both spoken and sung utterances (Brown et al., 2009). Indeed, several of the studies included in these meta-analyses reported activations in the ACC (Bach et al., 2008; Doherty et al., 2004; Gandour et al., 1998), and yet ALE foci in this region did not reach significance for any analysis. One likely explanation is that much of the ACC lies outside the volume covered in our "restricted" analyses and that our whole-brain analyses had too little power to detect convergence in this area. In addition, the anatomy of the ACC is variable across individuals (Paus et al., 1996), and it is therefore possible that differences between subjects in cingulate anatomy resulted in subtle variability in the localization of foci between studies.

Unlike the result in auditory areas, substantial divergence was observed in inferior frontal regions. More specifically, affective prosody activated the

IFG pars orbitalis (BA 47) bilaterally while linguistic prosody activated the IFG pars opercularis (ventral BA 44) bilaterally. The absence of overlap in the inferior frontal region suggests that activations here do not simply relate to some aspect of pitch processing alone, but may instead reflect the different classes of information that listeners extract from affective versus linguistic cues in speech prosody. Notably, Wildgruber et al., (2004) compared affective and linguistic prosody perception directly in an fMRI study. These authors observed a similar localization for both functions in inferior frontal regions.

We observed a small number of areas that were uniquely associated with each function of interest. The IFG pars orbitalis (BA 47), which was associated most strongly with affective prosody, is distinct from adjacent Broca's area in both cytoarchitecture (Brodmann, 1909/1994) and structural connectivity. The homologous region in macaques, area 47/12, receives projections from both limbic regions and the homologue of Broca's area (Petrides and Pandya, 2001). Area 47/12 is part of an orbitofrontal network that receives input from sensory areas, including auditory, visual, somatosensory, olfactory, visceral and gustatory cortices, as well as limbic areas such as the amygdala, subiculum, entorhinal cortex, and perirhinal cortex. This same network projects to the hypothalamus and periaquaductal grey by way of the ventromedial prefrontal cortex (Price, 1999). Diffusion tensor imaging in humans reveals a similar pattern. The IFG pars orbitalis is connected to auditory and visual areas via the inferior occipitofrontal fasciculus and middle longitudinal fasciculus (Turken and Dronkers, 2011). The frontal operculum adjacent to BA 47 is connected to the amygdala and septal region (Anwander et al., 2007)

and plays a role in emotion regulation in conjunction with the amygdalae and nucleus accumbens (Wager et al., 2008). Patients with lesions in this region and the adjacent orbitofrontal cortex have deficits in recognizing emotions in others as well as changes in behaviour and subjective emotional experience (Hornak, Rolls and Wade, 1996). This region is consistently active when subjects experience particular emotions or when they perceive emotions in either the auditory or visual domain (Lindquist et al., 2012). The IFG pars orbitalis may therefore be well situated to act as an interface between limbic and sensorimotor networks, as would be necessary for affective prosody perception. Indeed, given the diverse sensory information available to this region, it is not surprising that it is involved in the perception of emotional faces and gestures as well (Lotze et al., 2006; Sprengelmeyer et al., 1998).

Linguistic prosody, by contrast, was associated most strongly with the IFG pars opercularis (BA 44). Ventral BA 44 is associated with lexicosyntactic function. Functional MRI studies have shown that syntactic processing activates ventral IFG pars opercularis (Friederici et al., 2000; Heim et al., 2003a). In contrast, other linguistic functions, such as phonological processing, activate a locus in dorsal BA 44 (Heim et al., 2003b; Papoutsi et al., 2009). Given that linguistic prosody plays a role in syntactic disambiguation (Beach, 1991), it is perhaps not surprising that this suprasegmental element of speech shares brain areas with syntactic processing.

2.5.1 Lateralization versus localization

Both the neurological and neuroimaging literatures on the perception of prosody are concerned primarily with the lateralization of function in temporal and frontal language areas. In agreement with this literature, we observed consistent right-hemisphere lateralization in temporal-lobe auditory areas. Importantly, we observed this pattern of lateralization for both affective and linguistic prosodies, constituting a region of overlap between these functions. In contrast, our results did not support a consistent lateralization in the frontal lobe for either affective or linguistic prosody. Our meta-analyses instead demonstrated that bilateral inferior frontal activations were likely to be reported by neuroimaging studies of both functions, although in non-overlapping regions. While direct contrasts between conditions appeared to support the lateralization of affective prosody to the left inferior frontal gyrus, our primary analyses demonstrated that affective prosody perception did in fact activate right inferior frontal regions as well. Both affective and linguistic prosody activated bilateral (although distinct) inferior frontal regions, as demonstrated by Figure 1. However, this does not preclude the interpretation that some functional aspect of the task may influence patterns of lateralization. It has been proposed that one contributor to the frequent, but inconsistent, lateralization of speech prosody, especially in temporal-lobe auditory areas, is that the window of temporal integration of pitch information differs between the two hemispheres (Buchanan et al., 2000) such that the left hemisphere processes relatively fast frequency modulations and the right hemisphere relatively slow modulations (Zatorre, 2001).

2.5.2 Task-type

The middle frontal gyrus (part of the dorsolateral prefrontal cortex) and STG were associated with task-related differences in working-memory load in the linguistic-prosody meta-analysis. Activations in this region were more prominent in studies of linguistic prosody than affective prosody perception. This may be due to a greater proclivity towards experiments with high verbal working-memory demands in that literature. Studies of linguistic prosody used methods with either a high working memory load, namely same/different tasks, or with a low working memory load, namely forced-choice identification tasks. In comparison, studies of affective prosody used primarily tasks with low verbal working-memory demands. This methodological difference might account for the increased likelihood of observing activation in the middle frontal gyrus for linguistic prosody compared with affective prosody. This finding is corroborated by a meta-analysis that explicitly examined verbal working-memory demands (Chein et al., 2002). Note that this analysis was conducted to detect confounds in our primary analyses and should not be taken as an analysis of working memory per-se.

2.5.3 Production and perception

To the best of our knowledge, only one study has compared functional activations between perception and production of prosody, and it did so for both linguistic and affective prosody (Aziz-Zadeh et al., 2010). While that study did not observe activation in the IFG pars orbitalis that we described for affective prosody, it did observe activation in the left IFG pars opercularis for

the production and perception of both affective and linguistic prosodies. Our meta-analyses revealed ALE foci in this region, although the localization varied for each function. The IFG pars opercularis may be an important point of interaction for affective and linguistic prosody. More specifically, the IFG pars opecularis is purported to be a "mirror neuron" area involved in both the production and perception of actions (Aziz-Zadeh et al., 2006). This area may be structurally connected with the primary motor cortex (Greenlee et al., 2004; Simonyan et al., 2009). It may therefore constitute an area of convergence for affective and linguistic prosody en route to the motor cortex.

2.5.4 Prosody networks

A number of models have proposed temporo-frontal networks for prosody perception based on the activation patterns for affective prosody (Ethofer et al., 2006; Schirmer & Kotz, 2006). Two recent models have suggested that prosody perception occurs in three stages: 1) acoustic analysis in the voice-selective areas of the mSTS (Belin et al., 2000), 2) identification of vocally expressed emotion in the aSTG (Kotz & Paulmann, 2011) or pSTG (Brück et al., 2011) and 3) explicit evaluation of prosody in the IFG. These models agree with one another in most respects, with the exception of the localization of temporal regions specific for affective voices. While another prosody meta-analysis (Witteman et al., 2012) supported the localization of Brück et al. (2012) to the pSTG, our results supported the role of both aSTG and pSTG in affective prosody processing. Notably, we observed ALE foci in bilateral aSTG for affective prosody only, not for linguistic prosody. However these foci did not

survive a direct contrast between the two functions. Due to the lower power of the linguistic-prosody analysis relative to affective prosody, it cannot be concluded from the data that either the aSTG or pSTG is specific to affective prosody.

Conjunction analysis revealed several areas of common activation between affective and linguistic prosody. Among these were the right auditory association cortex, which is specialized for the fine-grained analysis of pitch (Zatorre & Gandour, 2008), left anterior insula, which is anatomically connected to the entire extent of the IFG (spanning the pars opercularis, pars triangularis and pars orbitalis; Catani et al., 2012), and the somatotopic oro-laryngeal portion of the SMA (Fried et al., 1991). This group of regions is likely involved in audio-vocal functioning generally, rather than prosody specifically.

Affective and linguistic prosody do not generally occur in isolation but rather in parallel with speech. A focus of future research should be to further develop network models of prosody perception and to extend these models to incorporate production with the aim of integrating these networks with extant models of speech. For example, the "Directions into Velocities and Articulators" (DIVA) model (Golfinopoulos et al., 2010) is a well established model of speech production that describes how intended speech sounds are converted into articulatory movements that ultimately result in the production of speech. Such a set of mechanisms should, in theory, accommodate the production of the pitch-based cues that are used for affective and linguistic prosodies.

The IFG pars opercularis locus observed for linguistic prosody is part of

Broca's area (and Broca's homologue) and is therefore already a component of most neural models of speech. However, the expression of emotion is acoustically similar whether it occurs without language in the form of affect bursts such as laughter and crying (Schröder, 2003) or with language in the form of affective prosody (Banse and Sherer, 1996). Affective prosody may therefore require the integration of an evolutionarily ancestral subcortical system for affective communication found in monkeys (Jürgens, 2009) with the evolutionarily recent cortical system for speech and language that is found only in humans. We suggest that the IFG pars orbitalis (BA 47) may function as such an interface between emotion and vocalization, although others have proposed that the ACC serves this function as well (Jürgens, 2009). One caveat to this proposal is the suggested role of the IFG pars orbitalis in other functions. This region has previously been reported in neuroimaging studies of both linguistic (Fiez, 1997) and musical semantics (Levitin and Menon 2003) as well as in pitch memory (Zatorre et al., 1994). Price (1999) noted that the orbital region of the macaque, including BA 47/12, is cytoarchitectonically diverse. Further research is needed to search for potential functional subdivisions within this region.

2.6 Limitations

A potential limitation of our analysis is that our dataset included more studies of affective prosody than linguistic prosody. This unbalanced design may have introduced some bias into the data and limited the inferences that could be made from it. We attempted to mitigate this limitation by checking the

number of studies that contribute to each of the foci in our contrasts.

Our analysis of working memory load relied on a small and unbalanced sample of studies of linguistic prosody perception. Furthermore, our division into high and low working memory load was confounded with the distinction between task-driven effects and stimulus-driven effects discussed by Witteman et al. (2012). For these reasons, we stress that our working memory results are provisional and are intended only to aid in the interpretation of the other analyses.

2.7 Conclusion

We meta-analyzed the literatures on the neural correlates of two pitch-based paralinguistic functions. The results provide mixed support for hemispheric lateralization of speech prosody, with greater lateralization seen in temporal-lobe auditory areas than in frontal-lobe evaluative areas. Instead, the results support a localizationist account based on differentiation of the two prosodic functions in the inferior frontal gyrus. Linguistic prosody is associated with a portion of the IFG pars opercularis that is involved in syntactic processing. Affective prosody is associated with the IFG pars orbitalis, which is connected with both limbic and speech-motor areas, making it a good candidate as an interface between emotion and voice.

2.8 Acknowledgements

This work was funded by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to S.B (grant no. 371336).

2.9 References

Anwander, A., Tittgemeyer, M., von Cramon, D.Y., Friederici, A.D., & Knösche, T.R. (2007). Connectivity-based parcellation of Broca's area. Cerebral Cortex, 17(4), 816−25.

Aziz-Zadeh, L., Koski, L., Zaidel, E., Mazziotta, J., & Iacoboni, M. (2006). Lateralization of the human mirror neuron system. Journal of Neuroscience, 26(11), 2964−70.

Aziz-Zadeh, L., Sheng, T., & Gheytanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. PLoS One, 5(1), e8759.

Bach, D.R., Grandjean, D., Sander, D., Herdener, M., Strik, W.K., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. Neuroimage, 42(2), 919−27.

Banse, R., & Sherer, K. (1996). Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology, 70(3), 614−36.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. Journal of Memory and Language, 30(6), 644−63.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P, & Pike, B. (2000) Voice-selective areas in human auditory cortex. Nature, 6767, 309−12.

Brodmann, K. (1909/1994). Localisation in the cerebral cortex (3rd ed.). Garey, L. J. (Ed.). New York, NY: Springer.

Brown, S., Laird, A.R., Pfordresher, P.Q., Thelen, S.M., Turkeltaub, P., & Liotti, M. (2009). The somatotopy of speech: Phonation and articulation in the human motor cortex. Brain and Cognition, 70(1), 31−41.

Brown, S., Martinez, M., & Parsons, L. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. Neuroreport, 13, 2033−7.

Brück, C., Kreifelts, B., Kaza, E., Lotze, M., & Wildgruber, D. (2011). Impact of personality on the cerebral processing of emotional prosody. Neuroimage, 58(1), 259−68.

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. Physics of Life Reviews, 8, 383−403.

Buchanan, T.W., Lutz, K., Mirzazade, S., Specht, K., Shah, N.J., Zilles, K. et al., (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. Cognitive Brain Research, 9(3), 227−38.

Catani, M., Dell'acqua, F., Vergani, F., Malik, F., Hodge, H., Roy, P., et al. (2012). Short frontal lobe connections of the human brain. Cortex, 48(2), 273−91.

Chein, J.M., Fissell, K., Jacobs, S., & Fiez, J.A. (2002). Functional heterogeneity within Broca's area during verbal working memory. Physiology and Behaviour, 77, 635−39.

de Mareüil, P.B., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. Phonetica, 63(4), 247−67.

Dietrich, S., Hertrich, I., Alter, K., Ischebeck, A., & Ackermann, H. (2008). Understanding the emotional expression of verbal interjections: A functional MRI study. Neuroreport, 19(18), 1751−5.

Doherty, C.P., West, W.C., Dilley, L.C., Shattuck-Hufnagel, S., & Caplan, D. (2004). Question/statement judgments: An fMRI study of intonation processing. Human Brain Mapping, 23(2), 85−98.

Eickhoff, S.B., Bzdok, D., Laird, A.R., Kurth, F., & Fox, P.T. (2012). Activation likelihood estimation meta-analysis revisited. Neuroimage, 59(3),

2349−61.

Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., & Fox, P.T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. Human Brain Mapping, 30(9), 2907−26.

Ekman, P., Sorenson, E.R., & Friesen, W.V. (1969). Pan-cultural elements in facial displays of emotion. Science, 164(3875), 86−8.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J. et al. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. Neuroimage, 30(2), 580−7.

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. Cerebral Cortex, 22(1), 191−200.

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P. et al., (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. Journal of Cognitive Neuroscience, 21(7), 1255−68.

Fairbanks, G. & Provonost, W. (1938). Vocal pitch during simulated emotion. Science, 88, 382−3.

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In: H. Papouěk, U. Jürgens, M. Papouěk (eds). Nonverbal vocal communication: Comparative and developmental approaches. New York: Cambridge University Press, p. 262−82.

Fiez, J. (1997). Phonology, semantics, and the role of the left inferior prefrontal cortex. Human Brain Mapping, 5(2), 79−83.
Fontaine, D., Capelle, L., & Duffau, H. (2002). Somatotopy of the supplementary motor area: Evidence from correlation of the extent of surgical resection with the clinical patterns of deficit. Neurosurgery, 50(2), 297−303.

Fried, I., Katz, A., McCarthy, G., Sass, K.J., Williamson, P., Spencer, S.S. et al., (1991). Functional organization of human supplementary motor cortex

studied by electrical stimulation. Journal of Neuroscience, 11(11), 3656−66.

Friederici, A.D., Meyer, M., & von Cramon, D.Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. Brain and Language, 75(3), 465−77.

Frühholz, S., Ceravolo, L., & Grandjean, D. (2011). Specific brain networks during explicit and implicit decoding of emotional prosody. Cerebral Cortex, 22(5), 1107−17.

Gandour, J., Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. Neuroreport, 9(9), 2115−19.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L. et al., (2003a). Temporal integration of speech prosody is shaped by language experience: An fMRI study. Brain and Language, 84(3), 318−36.

Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., & Li, X. (2003b). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. Human Brain Mapping, 18(3), 149−57.

Gay, T. (1978). Physiological and acoustic correlates of perceived stress. Language and Speech, 21(4), 347−53.

Golfinopoulos, E, Tourville, J., & Guenther, F. (2010). The integration of large-scale neural networks modeling and functional brain imaging in speech motor control. Neuroimage, 50(3), 862−74.

Gorelick, P.B., & Ross, E.D. (1987). The aprosodias: Further functional-anatomical evidence for the organisation of affective language in the right hemisphere. Journal of Neurology and Psychiatry, 50(5), 553−60.

Greenlee, J.D.W., Oya, H., Kawasaki, H., Volkov, I.O., Kaufman, O.P., Kovach, C. et al. (2004). A functional connection between inferior frontal gyrus and orofacial motor cortex in human. Journal of Neurophysiology, 92(2), 1153−64.

Heim, S., Opitz, B., & Friederici, A.D. (2003a). Distributed cortical networks for syntax processing: Broca's area as the common denominator. Brain and Language, 85(3), 402−8.

Heim, S., Opitz, B., Müller, K., & Friederici, A.D. (2003b). Phonological processing during language production: fMRI evidence for a shared production-comprehension network. Cognitive Brain Research, 16(2), 285−96.

Hoekert, M., Vingerhoets, G., & Aleman, A. (2010). Results of a pilot study on the involvement of bilateral inferior frontal gyri in emotional prosody perception: An rTMS study. BMC Neuroscience, 11, 1−8.

Hornak, J., Rolls. E., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. Neuropsychologia, 34, 247−61.

Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. Human Brain Mapping, 26(2), 128−38.

Johnstone, T., van Reekum, C.M., Oakes, T.R., & Davidson, R.J. (2006). The voice of emotion: An fMRI study of neural responses to angry and happy vocal expressions. Social Cognitive and Affective Neuroscience, 1(3), 242−29.

Jusczyk, P.W., Hirsh-Pasek, K., Nelson, D.G., Kennedy, L.J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. Cognitive Psychology, 24(2), 252−93.

Jürgens, U. (2002). Neural pathways underlying vocal control. Neuroscience & Biobehavioral Review, 26(2), 235−58.

Jürgens, U. (2009). The neural control of vocalization in mammals: A review. Journal of Voice, 23(1), 1-10.

Kotz, S., & Silke, P. (2011). Emotion, language, brain. Language and Linguistic Compass, 5, 108−125.

Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? Journal of Phonetics, 25(3), 313−42.

Levitin, D., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: A possible role for Brodmann area 47 in temporal coherence. Neuroimage, 20(4), 2142−52.

Lindquist, K., Wager, T.D., Kober, H., Bliss-Moreau, E., & Barrett, L.F. The brain basis of emotion: A meta-analytic review. Behavioral and Brain Sciences, 35, 121−43.

Lotze, M., Heymans, U., Birbaumer, N., Veit, R., Erb, M., Flor, H. et al. (2006). Differential cerebral activation during observation of expressive gestures and motor acts. Neuropsychologia, 44(10), 1787−95.

Juslin, P. & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? Psychological Bulletin, 129, 770−814.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., & von Cramon, D.Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Human Brain Mapping, 17(2), 73−88.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A.D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. Brain and Language, 89(2), 277−89.

Monrad-Krohn, G.H. (1947). The prosodic quality of speech and its disorders. Acta Psychiatrica Scandinavica, 3-4, 255−69.

Nagaratnam N., Nagaratnam, K., Ng, K., & Diu, P. (2004). Akinetic mutism following stroke. Journal of Clinical Neuroscience, 11(1), 25−30.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. Neuroimage, 25(3), 653−60.

Pearsall, J., Hanks, P., Soanes, C., & Stevenson, A (Eds.). (2005). Oxford English dictionary (2nd ed., revised). New York: Oxford University Press.

Papoutsi, M., de Zwart, J.A., Jansma, J.M., Pickering, M.J., Bednar, J.A., & Horwitz, B. (2009). From phonemes to articulatory codes: An fMRI study of the role of Broca's area in speech production. Cerebral Cortex, 19(9), 2156−65.

Paus, T., Tomaiuolo, F., Otaky, N., Petrides, M., Atlas, J., Morris, R. et al.,

(1996). Human cingulate and paracingulate sulci: Pattern, variability, asymmetry, and probabilistic map. Cerebral Cortex, 6, 207−14.

Pell, M.D. (1998). Recognition of prosody following unilateral brain lesion: Influence of functional and structural attributes of prosodic contours. Neuropsychologia, 36(8), 701−15.

Pell, M.D., & Baum, S.R. (1997). The ability to perceive and comprehend intonation in linguistic and affective contexts by brain-damaged adults. Brain and Language, 57(1), 80−99.

Petrides, M., & Pandya, D.N. (2001). Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. Neuroscience, 16, 291−310.

Price, J.L. (1999). Prefrontal cortical networks related to visceral function and mood. Annals of the New York Academy of Sciences, 877, 383−96.

Ross, E.D. (1981). The aprosodias: Functional-anatomic organization of the affective components of language in the right hemisphere. Archives of Neurology, 38(9), 561−69.

Sauter, D.A., Eisner, F., Ekman, P., & Scott, S.K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. Proceedings of the National Academy of Sciences, 107(6), 2408−12.

Sauter, D.A., & Scott, S.K. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? Motivation and Emotion, 31(3), 192−9.

Scherer, K., Banse, R., & Wallbott, H. (2001). Emotion inferences from vocal expression correlate across languages and cultures. Journal of Cross Cultural Psychology, 32(1), 76−92.

Schirmer, A., & Kotz, S.A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. Trends in Cognitive Science, 10(1), 24−30.

Schröder, M. (2003). Experimental study of affect bursts. Speech Communication, 40, 99−116.

Simonyan, K., Ostuni, J., Ludlow, C.L., & Horwitz, B. (2009). Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. Journal of Neuroscience, 29(47), 14912−23.

Sprengelmeyer, R., Rausch, M., Eysel, U.T., & Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. Philosophical Transactions of the Royal Society B, 265(1409), 1927−31.

Strelnikov, K.N., Vorobyev, V.A., Chernigovskaya, T.V., & Medvedev, S.V. (2006). Prosodic clues to syntactic processing: A PET and ERP study. Neuroimage, 29(4), 1127−34.

Turkeltaub, P.E., Eden, G.F., Jones, K.M., & Zeffiro, T.A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. Neuroimage, 16(3), 765−80.

Turkeltaub, P.E., Eickhoff, S.B., Laird, A.R., Fox, M., Wiener, M., & Fox, P. (2011). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. Human Brain Mapping, 33, 1−13.

Wager, T.D., Davidson, M.L. , Hughes, B.L., Lindquist, M.A., & Ochsner, K.N. (2008). Prefrontal-subcortical pathways mediating successful emotion regulation. Neuron, 59, 1037−50.

Weintraub, S., Mesulam, M., & Kramer, L. (1981). Disturbances in prosody: A right-hemisphere contribution to language. Archives of Neurology, 38, 742−44.

Wennerstrom, A. (2001). The music of everyday speech: Prosody and discourse analysis. New York: Oxford University Press. Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W. et al., (2008). Cerebral processing of emotional prosody: Influence of acoustic parameters and arousal. Neuroimage, 39(2), 885−93.

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders,S., Grodd, W., et al., (2004). Distinct frontal regions subserve evaluation of linguistic and affective aspects of intonation. Cerebral Cortex, 14I, 1384−9.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T. et

al. (2005). Identification of emotional intonation evaluated by fMRI. Neuroimage, 24(4), 1233−41.

Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. International Journal of Speech-Language Pathology, 11, 277−281.

Witteman, J., van Heuven, V.J.P., & Schiller, N. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. Neuropsychologia, 50, 2752−63.

Witteman, J., van Ijzendoorn, M., van de Velde, D., van Heuven, V.J.P., & Schiler, N.O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. Neuropsychologia, 49, 3722−38.

Xu, Y., & Xu, C.X. (2005). Phonetic realization of focus in English declarative intonation. Journal of Phonetics, 33(2), 159−97.

Zatorre, R.J. (2001). Neural specializations for tonal processing. Annals of the New York Academy of Sciences, 930, 193−210.

Zatorre, R.J., Evans, A.C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. Journal of Neuroscience, 14(4), 1908−19.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 363(1493), 1087−104.

Chapter 3

Pitch underlies activation of the vocal system during affective vocalization

Michel Belyk & Steven Brown.

Social Cognitive and Affective Neuroscience, In Press.

3.1 Abstract

Affective prosody is that aspect of speech that conveys a speaker's emotional state through modulations in various vocal parameters, most prominently pitch. While a large body of research implicates the cingulate vocalization area in controlling affective vocalizations in monkeys, no systematic test of functional homology for this area has yet been reported in humans. In this study, we used functional magnetic resonance imaging to compare brain activations

when subjects produced affective vocalizations in the form of exclamations vs non-affective vocalizations with similar pitch contours. We also examined the perception of affective vocalizations by having participants make judgments about either the emotions being conveyed by recorded affective vocalizations or the pitch contours of the same vocalizations. Production of affective vocalizations and matched pitch contours activated a highly overlapping set of brain areas, including the larynx-phonation area of the primary motor cortex and a region of the anterior cingulate cortex that is consistent with the macro-anatomical position of the cingulate vocalization area. This overlap contradicts the dominant view that these areas form two distinct vocal pathways with dissociable functions. Instead, we propose that these brain areas are nodes in a single vocal network, with an emphasis on pitch modulation as a vehicle for affective expression.

Key words: affective prosody; fMRI; larynx-phonation area; cingulate cortex; voice; pitch Introduction

3.2 Introduction

Affective prosody − or tone of voice − conveys a speaker's emotional state through modulations in the acoustics of the voice, particularly vocal pitch (Banse and Sherer, 1996). Listeners can reliably recognize a broad range of vocally-expressed emotions, even when the spoken words are unrelated to the emotion (Belin et al., 2008; Fairbanks and Pronovost, 1938; Simon-Thomas et al., 2009) or when recordings are filtered to remove segmental content (Lieberman and Michaels, 1962). Unlike the words that make up the segmental aspect

of speech, affective vocalizations can be recognized across languages (Laukka et al., 2013), between cultures that have had only minimal historical contact (Sauter et al., 2010) − although with some cultural variation (Scherer and Wallbott, 1994) − and across species (Farag et al., 2014). Indeed, infants who are hearing-impaired produce affective vocalizations that are acoustically similar to those of normal-hearing infants (Scheiner et al., 2004, 2006).

While affective prosody usually unfolds across the course of an utterance, it can also be uttered in the form of short bursts. Schröder (2003) proposed that such bursts occur along a continuum from "raw affect bursts" to "verbal interjections". Raw affect bursts, such as bouts of laughter or crying, are most similar to the innate affective calls of non-human animals. On the other hand, verbal interjections, such as exclamations, can be comprised of affectively-intoned single syllables (Belyk and Brown, 2014c). Exclamations are particularly well-suited for neuroimaging studies of affective prosody since: 1) they are brief enough to accommodate event-related designs in functional magnetic resonance imaging (fMRI) experiments, 2) they do not require syntactic processing, and 3) many of them are non-words, thereby reducing the necessity of semantic processing.

Compared to the vocal systems of animals studied in neural models of vocalization, such as the squirrel monkey (Saimiri sciureus) and rhesus monkey (Macaca mulatta), relatively little is known about how affective vocalizations are produced and controlled in humans. The lower motor neurons that innervate the musculature of the larynx − which is the organ of vocalization − are contained in the nucleus ambiguus of the medulla. Two motor pathways

project to the nucleus ambiguus to drive vocalization. The first is a pathway from the cingulate vocalization area (Jürgens and Pratt, 1979b) in the anterior cingulate cortex (ACC) that projects to the periaqueductal gray (PAG) of the midbrain, which itself projects to the nucleus ambiguus (Jürgens and Pratt, 1979b). In monkeys, stimulation of the PAG elicits species-specific affective vocalizations (Jürgens and Ploog, 1970), and lesioning of the PAG abolishes affective vocal responses both to environmental stimuli and to stimulation of cortical areas that project to the PAG (Jürgens and Pratt, 1979a; Jürgens and Pratt, 1979b). One such area is the ACC, stimulation of which also elicits species-specific affective vocalizations (Jürgens and Pratt, 1979b). Lesions to the ACC prevent the initiation of operantly conditioned vocalizations (Aitken, 1981; Sutton et al., 1974, 1981), but have no effect on spontaneous vocalizations in affective contexts that would normally elicit these responses (Jürgens and Pratt, 1979). Hence, the ACC in monkeys is believed to initiate volitional, but not reflexive, species-specific affective vocalizations via projections to the PAG (Jürgens, 2002, 2009).

A second pathway to the nucleus ambiguus originates in the cortical larynx area of the primary motor cortex of the precentral gyrus. In monkeys, electrical stimulation of the cortical larynx area stimulates contraction of the intrinsic and extrinsic laryngeal muscles (Hast et al., 1974), but does not elicit vocalization (Jürgens, 1974). Furthermore, bilateral lesions to this region have little effect on vocal behaviour (Jürgens et al., 1982; Kirzinger and Jürgens, 1982). Hence, the monkey larynx area, while clearly a source of innervation of

the laryngeal muscles, does not appear to drive vocal behaviour. This represents a striking species difference in the cortical control of the larynx between monkeys and humans, one supported by neuroanatomy. In monkeys, the cortical larynx area is restricted to the premotor cortex (Hast et al., 1974), while in the human brain it extends into primary motor cortex as well (Belyk and Brown, 2014b; Brown et al., 2008; Loucks et al., 2007; Simonyan et al., 2009). The human larynx area is activated by volitional movement of the laryngeal muscles, phonation, and forced expiration, leading us to refer to it as the "larynx phonation area" (LPA; Brown et al., 2008). In fact, in contrast to the monkey cortical larynx area, electrical stimulation of the human LPA does indeed produce vocalization (Penfield and Boldrey, 1937), and lesions may cause mutism (Jürgens et al., 1982) and other speech disorders. Furthermore, the human LPA projects monosynaptically to the laryngeal lower motor neurons in the nucleus ambiguus (Iwatsubo et al., 1990; Kuypers, 1958), while the monkey larynx area projects to the nucleus ambiguus only indirectly via synapses in the reticular formation (Jürgens and Ehrenreich, 2007; Simonyan and Jürgens, 2003).

One prominent model of vocal-motor control (Ackermann et al., 2014; Myers, 1976; Owren et al., 2011) dichotomizes innate-affective and learned non-affective vocal behaviours, attributing the production of affective vocalizations to the cingulate pathway and non-affective vocalizations to the primary motor cortex. Evidence from the monkey vocal system provides strong support for the involvement of the cingulate vocal pathway in driving species-specific affective vocalizations (Jürgens and Pratt, 1979; Jürgens and Pratt, 1979;

Kirzinger and Jürgens, 1982; Sutton et al., 1981). However, monkeys are not vocal learners (Petkov and Jarvis, 2012); innate, species-specific affective vocalizations constitute their entire vocal repertoire. In addition, the cortical larynx area of the monkey appears to lack vocal functionality altogether. This makes the monkey a poor model of the vocal-motor control of learned vocalizations. Humans clearly possess both of these vocal-motor pathways, but it is unclear whether these pathways operate reciprocally, as implied in this model, or if they operate in concert to produce the complex vocal repertoire of humans, which includes not only affective vocalizations but speech and song as well.

Given the anatomical and functional reorganization of laryngeal motor control in humans, it is unclear whether the LPA plays as limited a role in human affective vocalization as the cortical larynx area does in monkeys. Indeed, much of human affective vocal expression occurs in parallel with speech. Barrett et al. (2004) collected fMRI data while participants performed a speech task before and after inducing sad affect. Both self-reported sad affect and reduced fundamental-frequency range − a vocal cue of sadness − were correlated with the degree of activation of the ACC, as predicted from knowledge of the monkey vocal system. Similarly, Wattendorf et al. (2013), in an fMRI study of laughter, showed that both the ACC and PAG were active when laughter was voluntary, whereas inhibition of laughter activated the ACC without the PAG, and induced laughter activated the PAG without the ACC. These data are consistent with the roles of the PAG in producing affective vocalizations and the ACC in exerting volitional control over the PAG. However, activation

of the LPA also correlated with reduced fundamental-frequency range during induced sad affect (Barrett et al., 2004), and the LPA was active during both volitional and induced laughter (Wattendorf et al., 2013). These findings suggest that, in humans, control of the larynx during affective vocalization may result from an integration of the cingulate and primary-motor vocalization pathways.

We previously hypothesized (Belyk and Brown, 2014a) that a portion of the inferior frontal gyrus − the IFG pars orbitalis (IFGorb) − that is reliably activated when perceiving affective vocalizations may be involved in planning affective vocalizations. This region has anatomical connections with limbic, auditory, and premotor areas (Anwander et al., 2007; Price, 1999; Turken and Dronkers, 2011; Yeterian et al., 2012) similar to adjacent Broca's area, which sits at the interface of speech perception and production (Watkins and Paus, 2004). None of the handful of studies that have examined the production of affective vocalizations in humans have reported activation in the IFGorb (Aziz-Zadeh et al., 2010; Barrett et al., 2004; Wattendorf et al., 2013). However, given the relatively low power of whole-brain analyses in human brain imaging, we sought to perform a sensitive test of the activation of the IFGorb during affective production using a region-of-interest (ROI) analysis.

Because of the great paucity of research on affective vocalizing in humans, we conducted a highly controlled study that compared the production of exclamations to acoustically-matched nonsense syllables that were similar in vocal dynamics − particularly pitch profiles − but that differed in affective content.

In doing so, we aimed to test two hypotheses. First, in order to test for a potential dissociation between the cingulate and LPA vocal pathways, we examined whether these cortical regions were differentially activated during affective and non-affective vocal production. The putative dissociation between vocal-motor pathways for affective and non-affective vocalizations predicts that the ACC and PAG are more active when producing affective vocalizations and that the LPA is more active when producing non-affective vocalizations. Such a dissociation further predicts functional connectivity between the ACC and PAG, but not between the LPA and either of these structures. Second, given the acknowledged role of the IFGorb in perceiving affective vocalizations, we wanted to see if this area was active when producing affective vocalizations as well. In order to localize this area, we examined patterns of brain activation when participants made judgments about the emotions expressed in vocal recordings, as compared to judgments about the pitch contours of these recordings. We then performed ROI analyses to test the hypothesis that the IFGorb is also activated beyond baseline when producing affective vocalizations. Finally, given that the majority of the literature on affective prosody focuses on perception alone, we were interested in comparing the brain network for affective vocalizing with that for affective perception. Considering the general propensity of sensorimotor systems to activate during both action execution and observation (Aziz-Zadeh et al., 2010; Cross, Hamilton, & Grafton, 2006; Menenti, Gierhan, Segaert, & Hagoort, 2011), we performed exploratory analyses to search for exclusivity and overlap in our vocal production and perceptual experiments.

3.3 Methods

3.3.1 Stimulus recording procedure

A male professional actor was instructed to vocalize the monosyllables eep, ep, oap, and oop (phonetically /ip/, /ɛp/, /əp/, and /up/ in the International Phonetic Alphabet), in each of four different affective prosodies: happiness, sadness, pleasure, and disgust. He was given the instruction to produce happy and sad vocalizations with a descending pitch-contour, and to produce pleasure and disgust vocalizations with an arched pitch-contour. Four renditions of each syllable/emotion combination were recorded, resulting in a total of 64 recordings. The amplitude of the set of recordings was equalized in Praat (Praat: doing phonetics by computer; www.fon.hum.uva.nl/praat/) to ensure that all recordings were equally audible in the MRI environment.

3.3.2 Stimulus validation procedure

Twelve participants were presented with each of the 64 recordings, and performed a four-alternative forced-choice discrimination task. Participants were instructed to identify the emotion that was being expressed among the four possible responses of "happiness, "sadness", "pleasure", and "disgust". Recordings of syllable/emotion combinations were then ranked with regard to their accuracy of discrimination. For each emotion, recordings in the top two quartiles were included in the perceptual tasks of the fMRI experiment, while recordings in the third quartile were reserved for training prior to the scanning session. The remaining recordings were discarded. All subjects provided

written informed consent before participation in the study, which was approved by the McMaster Research Ethics Board.

3.3.3 Imaging experiment procedure

Sixteen participants with no history of neurological or psychiatric illness were recruited for the fMRI experiment. The data from two participants was omitted from the group analysis due to excessive head motion, resulting in a final sample size of 14 (10 female, 13 right handed). All subjects provided written informed consent before participation in the study, which was approved by the Hamilton Integrated Research Ethics Board.

Participants performed four tasks in four separate functional scans (i.e., one task per scan) in counterbalanced order, each scan lasting 7 minutes 40.8s. Each scan consisted of 32 trials lasting 14.4s (nine functional volumes) each. For the two vocal production tasks (see below), each trial consisted of a visual cue presented for 1.6s that prompted participants to vocalize, followed by 12.8s of fixation on a crosshair. For the two perception tasks (see below), each trial consisted of 1.6s of a recorded vocalization, followed by 12.8s of silence with visual fixation on a crosshair throughout. All tasks were modeled according to a slow event-related design.

Affective vocal production

Participants were visually cued with a monosyllabic exclamation-word, and were instructed to produce the exclamation with the appropriate prosody as expressively as possible. The exclamation-words were drawn from a previous

study in which participants expressively produced these tokens on command (Belyk and Brown, 2014c). The exclamation words were "Oooh!", "Mmm!", "Eww!", "Yuck!", "Yay!", "Good!", "No!" and "Damn!". Participants were trained on a day prior to the scanning session to perform the task while lying still and without making facial expressions or head movements.

Pitch-contour production

Participants produced monosyllables with either descending or arched pitch contours. The goal was to create a vocal task in which the acoustic properties of the exclamations were controlled for, but which lacked their affective character. On each trial, participants were visually cued with one of the monosyllables "eep", "ep", "oap" or "oop" as well as a pitch-contour representation in the form of either a downward-sloping arrow or an arch-shaped arrow. Hence, participants produced eight distinct vocalizations to match the number of both the exclamation words in the affective vocal-production condition and the vocal recordings in the perceptual tasks. These two contours were selected to match those of the vocal recordings in the perception task, which themselves were either descending or arched. Participants were trained on a day prior to the scanning session to perform this task while lying still and without making facial expressions or head movements.

Affect perception

Participants listened to the 32 recordings selected from the stimulus-validation experiment (see above) using MRI-compatible, noise-cancelling headphones.

Unlike the stimulus-validation study − where participants had to perform an emotion-identification task − the participants in the fMRI version of the task performed a binary valence-discrimination task. Participants were instructed to discriminate between recordings that expressed an emotion of either positive valence (i.e., happiness or pleasure) or negative valence (i.e., sadness or disgust). The assignment of valence to response buttons (index finger vs. middle finger) was counterbalanced across participants. In addition, the order of stimuli within each condition was randomized.

Pitch-contour perception

Participants listened to the same 32 recordings as in the affect perception task but this time performed a binary contour-discrimination task. Participants were instructed to discriminate between recordings that contained either a descending contour (i.e., happiness and sadness) or an arched contour (i.e., pleasure and disgust).

3.3.4 Magnetic resonance imaging

Magnetic resonance images were acquired with a GE Signa Excite 3 Tesla MRI. Functional images sensitive to the blood oxygen level-dependent (BOLD) signal were collected with gradient echo sequences with repetition time = 1600 ms, echo time = 33 ms, flip angle = 90 degrees, 28 slices, slice thickness = 4 mm, gap = 0 mm, in-plane resolution 3.75 mm by 3.75 mm, matrix = 64 x 64, and field of view = 240 mm. A total of 293 volumes was collected per scan. Five dummy volumes were discarded at the beginning of each scan, leaving a

total of 288 volumes per scan.

3.3.5 Image analysis

Functional scans were analyzed with Brain Voyager 2.4, supplemented with NeuroElf (neuroelf.net). Each functional scan was spatially smoothed with a Gaussian kernel of 4mm full-width-half-maximum and high-pass filtered with a cut-off frequency of 0.0078125 Hz (or 1/128s). Each sample was realigned to the first sample in order to correct for head motion. Head-motion correction generated a set of parameters indicating the extent of translation in and rotation around the three cardinal axes for each sample. These motion parameters were included as nuisance regressors in all subsequent analyses. Low-level contrasts were thresholded with a False Discovery Rate (FDR) of $p<0.005$ and a cluster threshold $k>24$. The cluster threshold was relaxed to $k>5$ in the midbrain to permit the detection of small nuclei in that region. High-level contrasts were thresholded at $p<0.01$ uncorrected. To control for the rate of false positives, a Monte Carlo simulation using the AlphaSim algorithm selected a cluster-size threshold for each high-level contrast that maintained a family-wise error rate of $p<0.05$.

Region-of-interest (ROI) analyses were conducted to determine 1) whether the cingulate and primary-motor vocal pathways are dissociated in function, and 2) whether activation of the IFG pars orbitalis is specific to making judgments about affective prosody or if it participates in encoding it during vocal

production as well. Five-millimeter cubic ROIs were placed around functionally defined maxima in the left ACC and PAG, bilateral LPA, and right IFGorb. Only unilateral ROIs were analyzed for the ACC, PAG, and IFGorb because full-brain analyses failed to localize equivalent contralateral regions. Regression coefficients from each ROI were entered into linear mixed models in R (R Core Team, 2014) with the crossed factors Task (production vs. perception) and Content (affect vs. pitch contour).

Functional connectivity analyses were conducted to test 1) whether activations in the ACC and PAG are differentially correlated when producing affective vocalizations versus matched pitch-contours, and 2) whether activations in the ACC and LPA are correlated, anti-correlated, or independent. Time courses were extracted from each ROI for all participants. First-level regression models were computed in R in which the time course of the PAG or LPA were predicted by the time course of the ACC, with the hemodynamic response function of the experimental design as a covariate. Notably, this analysis removes the mutual influence of the experimental design on the time course of each region, precluding the interpretation that these regions are merely co-activated. Correlation coefficients from the first-level analysis were tested in a second-level analysis using Welch's t-tests for samples of unequal variance in order to determine whether the regression coefficients differed significantly from zero and/or between the two production tasks.

3.4 Results

3.4.1 Behavioural performance

During debriefing, all participants reported successfully producing the target affective vocalizations and pitch contours on cue. All participants performed above chance level on both the valence (mean 93%, SD 0.06%) and pitch-contour discrimination tasks (mean 87%, SD 0.07%). While participants performed above chance on both tasks, their discrimination accuracy was higher for the valence task than the contour-discrimination task $t(10) = 3.84$, $p < 0.05$.

3.4.2 Production

Overlapping vocal production activations

The affective production task and the acoustically-matched pitch-contour production task activated a highly overlapping set of brain regions when compared to rest. These brain regions described a basic vocal network (see Figure 1A and Table 1). Both tasks activated the bilateral larynx phonation area and adjacent orofacial motor cortex (BA 4), extending into the premotor cortex (lateral BA 6) and primary somatosensory cortex (BA 3/1/2). Bilateral activation was also observed in the SMA (medial BA 6) extending into the ACC (BA 24), anterior insula, thalamus, primary auditory cortex (BA 41), and visual cortex (spanning BA 17, 18 and 19, and reflecting the presence of the visual stimulus), left STG (BA 22), as well as the right superior parietal lobule (SPL) and putamen.

Non-overlapping vocal production activations

Producing affective vocalizations further activated the left parahippocampal gyrus (BA 19) and globus pallidus, and right subthalamic nucleus, cerebellum, and PAG. Producing meaningless pitch contours further activated the right caudate nucleus and left posterior gyrus (BA 23).

Producing affective vs pitch contour vocalizations

Contrary to the predicted dissociation, there were no significant differences in brain activation between the production of affective vocalizations and matched pitch-contours anywhere in the brain.

Table 3.1: Production. Locations of peak voxels for producing affective and contour-based vocalizations. After each anatomical name in the brain region column, the Brodmann number for that region is listed. The columns labeled as x, y, and z contain the Talairach coordinates for the peak of each cluster reaching significance at FDR p>0.005 with cluster threshold k>24. ACC: anterior cingulate cortex; IFG: inferior frontal gyrus LPA: larynx-phonation area; PAG: periaqueductal grey; PrhG: parahippocampal gyrus; SMA: supplementary motor area; STG: superior temporal gyrus; SPL: superior parietal lobule; S1: primary somatosensory cortex.

	Affective Production					Contour Production				
Brain Region	x	y	z	Voxels	t-value	x	y	z	Voxels	t-value
Frontal Lobe										
LPA/S1 (BA 4/6/3)	39	-16	34	570	11.5	54	-13	43	49	8.7
	-57	-10	37	366	9.7	-48	-7	34	31	7.1
SMA (BA 6)						3	-4	52	815	9.0
	-6	-16	55	401	9.9	-9	-10	61	40	8.7
Precentral gyrus (BA 6)						24	-16	52	48	7.0
						54	-7	28	855	11.6
ACC (BA 24)	-9	5	37	186	8.8					
IFG/insula	36	14	13	179	7.3	30	17	13	380	8.4
	-30	14	19	472	9.5	-33	20	13	29	6.9
Parietal Lobe										
SPL (BA 7)	-18	-34	61	32	7.5	-60	-16	28	659	9.2
	33	-58	46			15	-73	46	83	8.8
						24	-58	46	64	6.6
						18	-70	55	30	6.2
						-36	-61	46	143	7.3

Continued on next page

	Affective Production					Contour Production				
Brain Region	x	y	z	Voxels	t-value	x	y	z	Voxels	t-value
Posterior cingulate (BA 23)						-9	-25	28	27	6.7
Temporal Lobe										
Heschl's gryus (BA 41)	51	-22	10	298	9.7	45	-19	10	41	7.1
Heschl's gryus (BA 41)	-48	-31	10	451	10.2	-54	-28	10	205	8.8
PrhG (BA 19)	-18	-49	-2	298	9.2					
STG (BA 22)	-51	-13	10	25	8.0	-54	8	4	402	7.9
Occipital Lobe										
Lingual gyrus (BA 17)	-18	-94	-11	737	10.7	-27	-97	-2	52	7.5
Inferior occipital gyrus (BA 18)	-42	-82	-14	71	9.9	-39	-85	-5	1681	10.4
Inferior occipital gyrus (BA 19)						42	-79	-5	128	8.0
Fusiform gyrus (BA 19)	15	-64	1	332	8.5	21	-61	-5	1710	9.9
	-30	-88	1	36	9.9	-42	-76	-11	75	9.1
Middle occipital gyrus (BA 18)	18	-91	19	99	6.4	21	-91	13	124	8.4
Lingual gyrus (BA 18)	-15	-73	7	59	7.6	-12	-73	4	219	9.5
Lingual gyrus (BA 19)	-12	-85	25	86	7.6	-24	-79	28	203	8.3
Cuneus (BA 18)	24	-94	1	113	8.3	24	-79	22	101	6.9
Subcortical										
Thalamus	9	-25	1	318	10.1	6	-16	4	129	6.6
	-21	-25	1	354	12.5	-15	-22	-2	200	7.4
Subthalamic nucleus	15	-13	-2	23	9.0					
Caudate nucleus						12	8	13	77	7.7
Cerebellum	9	-55	-14	291	8.5					
Putamen	18	8	4	95	7.8	21	8	13	226	7.5
Globus palidus	-24	5	4	99	7.7					
Periaquaductal gray	6	-31	-8	8	5.2					

3.4.3 Perception

Overlapping perceptual activations

The two perceptual judgment tasks activated highly overlapping brain regions when each task was compared to rest. These regions included auditory, language, and motor areas related to the perceptual judgment, as well as visual areas related to the response cue (see Figure 1B and Table 2). Both tasks activated the bilateral primary auditory cortex (BA 41), visual cortex (BA's 17 and 18), thalamus, anterior insula (BA 13), and inferior parietal lobule (IPL; BA 40), right orofacial premotor cortex (BA 6), SMA (BA 6), cerebellum, claustrum, middle frontal gyrus (BA 9), and left posterior cingulate (BA

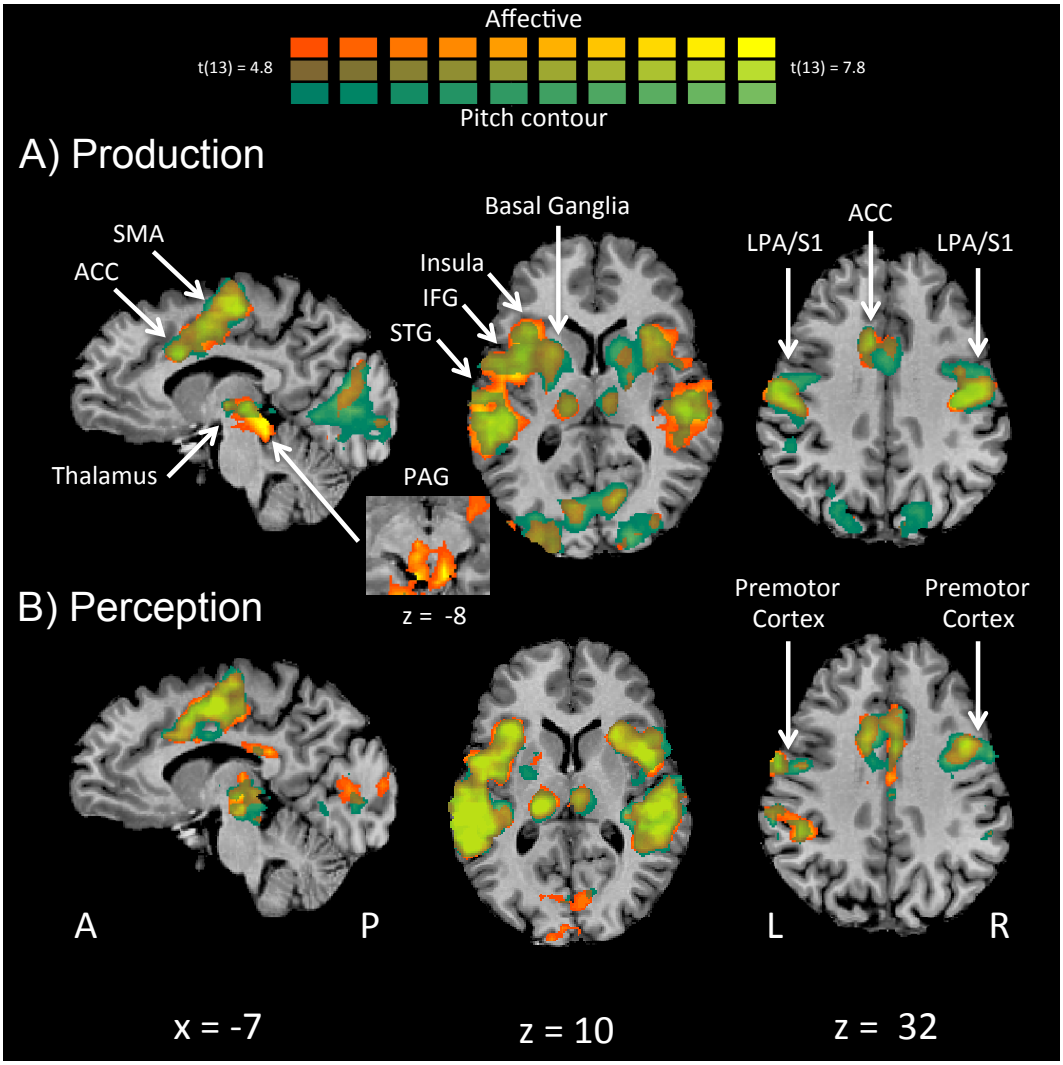

Figure 3.1: Whole brain analyses.

A) Affective vocal production (orange) and pitch-contour vocal production (green) versus rest reveal a remarkable degree of overlap across most of the vocal network, demonstrating little specificity for affective vocalization. B) Affect perception (orange) and pitch-contour perception (green) versus rest again demonstrate strong overlap between conditions. ACC: anterior cingulate cortex; IFG: inferior frontal gyrus LPA: larynx-phonation area; PAG: periaqueductal grey; SMA: supplementary motor area; STG: superior temporal gyrus; S1: primary somatosensory cortex.

23), precentral and postcentral gyri (BA 4/3), corresponding to the hand area involved in button press. Both conditions also activated the same region of the ACC observed during the vocal production tasks, although with a slightly reduced magnitude.

Non-overlapping perception activations

Making judgments about the emotions expressed in the vocal stimuli further activated the temporal pole, right IFG pars opercularis (BA 44), and IPL (BA 40), as well as the left IFG pars triangularis (BA 45), posterior cingulate gyrus (BA 23), putamen, and cerebellum. Making judgments about the pitch contour of the vocal stimuli further activated the bilateral pSTG (BA 22), left IFG pars opercularis (BA 44), superior parietal lobule (SPL; BA 7), and SMA, as well as the right orofacial premotor cortex (BA 6).

Table 3.2: Perception. Locations of peak voxels for the affect perception and pitch-contour perception tasks. After each anatomical name in the brain region column, the Brodmann number for that region is listed. The columns labeled as x, y, and z contain the Talairach coordinates for the peak of each cluster reaching significance at FDR p>0.005 with cluster threshold k>24. IFG: inferior frontal gyrus; IPL: inferior parietal lobule; ITG; inferior temporal gyrus; SMA: supplementary motor area; pSTG: posterior superior temporal gyrus; PCG: posterior cingulate gyrus SPL: superior parietal lobule.

	Affective Perception					Contour Perception				
Brain Region	x	y	z	Voxels	t-value	x	y	z	Voxels	t-value
Frontal Lobe										
SMA (BA 6)	6	2	46	1223	14.7	6	2	49	57	12.2
IFG pars opercularis (BA 44)	45	8	10	758	13.3					
						-45	5	16	511	12.6
IFG pars triangularis (BA 45)	-48	5	4	449	10.1					
Hand motor cortex (BA 4)						51	-10	46	81	8.8
	-51	-13	52	61	10.8	-54	-13	52	61	10.5
Premotor cortex (BA 6)	39	-10	49	94	7.6					
	-57	2	28	63	9.7	-9	-7	49	899	12.3
Anterior insula (BA 13)	-30	20	13	60	9.8	-33	11	13	76	10.7
Middle frontal gyrus (BA 9)	45	8	34	147	9.1	45	5	31	222	12.0

Continued on next page

	Affective Perception					Contour Perception				
Brain Region	x	y	z	Voxels	t-value	x	y	z	Voxels	t-value
						48	23	28	31	6.5
	-33	44	34	28	5.3					
IFG pars orbitalis (BA 47)		17	-8	34	8.4					
Temporal Lobe										
Heschle's gyrus (BA 41)	48	-28	10	805	11.6	57	-22	13	821	11.5
	-48	-34	10	1205	13.4	-54	-28	13	1148	13.1
pSTG (BA 22)						51	-40	13	36	9.4
						-57	-46	13	37	11.7
ITG (BA 47)						-57	-58	-5	33	6.3
Parietal Lobe										
IPL (BA 40)	48	-31	40	44	6.5					
	-33	-37	49	959	14.6	-60	-34	31	26	9.3
Postcentral gyrus (BA 3)						-30	-28	64	1145	15.0
PCG (BA 23)	6	-43	25	72	7.5	6	-31	28	97	9.3
	-9	-28	25	28	7.5					
SPL (BA 7)						-24	-70	43	61	8.3
Precuneus (BA 7)						-18	-82	46	28	6.4
Occipital Lobe										
Cuneus (BA 18)	6	-94	7	107	7.0					
	0	-91	16	29	6.4					
Lingual gyrus (BA 17)	6	-85	1	97	6.9	6	-73	4	155	7.4
	-15	-70	4	222	7.6	-15	-70	1	224	8.4
Subcortical										
Thalamus	3	-16	1	78	8.1	6	-16	10	65	8.2
						6	-4	1	26	7.7
	-18	-22	10	396	10.6	-15	-22	13	546	11.9
Cerebellum	15	-46	-17	358	8.8	15	-55	-17	350	10.3
	-18	-64	-26	81	8.2					
	-27	-61	-26	49	7.7					
Claustrum	33	-1	-2	34	7.4	30	11	13	823	14.9
Putamen	-21	5	7	41	5.5					

Perceiving affective vs pitch contour vocalizations

In replication of previous studies, contrasts between discriminating the emotion expressed in the stimuli versus discriminating the pitch contours that they contained revealed increased activation in the right anterior STG (BA 38), and left IFG pars triangularis (BA 45) and IPL (see Table 3).

Table 3.3: Affective vs pitch-contour tasks.

Locations of peak voxels for the high-level contrast [affective perception > contour perception]. Family-wise error was maintained at $p<0.05$ by combining an uncorrected threshold of $p<0.01$ with a cluster threshold of $k>12$, as selected by AlphaSim. No significant differences were observed for the contrast affect production > contour production. IFG: inferior frontal gyrus.

	Affect > Contour Perception				
Brain Region	x	y	z	Voxels	t-value
Frontal Lobe					
IFG pars triangularis (BA 45)	-46	25	2	43	4.2
Temporal Lobe					
Temporal Pole (BA 38)	42	17	-14	30	5.2
Parietal Lobe					
Inferior Parietal Lobule (BA 40)	-42	-37	55	23	5.0

3.4.4 Production vs perception

In order to compare the brain areas involved in production vs. perception irrespective of affective content, we ran the conjunction of contrasts [affect production > affect perception] ∩ [contour production > contour perception]. An examination of these two contrasts individually revealed nearly identical activation profiles. At a false discovery rate of $p<0.05$, only the bilateral LPA (extending into adjacent orofacial motor and somatosensory cortex), basal ganglia and visual cortex (reflecting the visual stimulus) were significantly more active during vocal production than the perceptual tasks. This small subset of activations reflects the otherwise overlapping activation of the vocal network for vocal production and perception tasks.

3.4.5 Region-of-interest analyses

Five regions of interest were defined around the functionally localized left ACC (Talairach coordinates -9, 5, 37), left PAG (-6, -31, -8), left (-52, -17, 33) and right (50, -11, 30) LPA/orofacial motor cortex, and right IFG pars orbitalis (42, 17, -8). A 2 x 2 Task (production vs. perception) by Content (affect vs. contour) analysis was conducted in each ROI. A significant main effect of task was observed for each ROI, with no effect of content and no task-by-content interaction (see Figure 2).

In the left ACC, activation was greater for the production tasks than the perception tasks $F(1,39)=14.4$, $p<0.05$. Neither content $F(1,39)=0.5$, $p=0.48$ nor the task-by-content interaction was significant $F(1,39)=0.014$, $p=0.97$. Similarly, activation in the PAG was greater for the production tasks than the perception tasks $F(1,39)=6.5$, $p<0.05$. Neither the main effect of content $F(1,39)=0.2$, $p=0.66$ nor the task-by-content interaction was significant $F(1,39)=0.46$, $p=0.5$. Since the PAG is of relatively small volume and a 5mm cubic ROI may extend beyond the PAG to include other midbrain nuclei, we replicated this analysis with 3mm and 1mm cubic ROIs. The results of these analyses were similar to those reported above, demonstrating that this analysis is robust to the size of the ROI (see Supplementary Table 2).

The left LPA was also more active for production tasks $F(1,39)=145$, $p<0.05$, with no effect of content $F(1,39)=0.8$, $p=0.38$ and no interaction $F(1,39)0$, $p=0.96$. Similar trends were observed in the right LPA with a significant main effect of task $F(1,39)=113$, $p<0.05$, no effect of content $F(1,39)=1$,

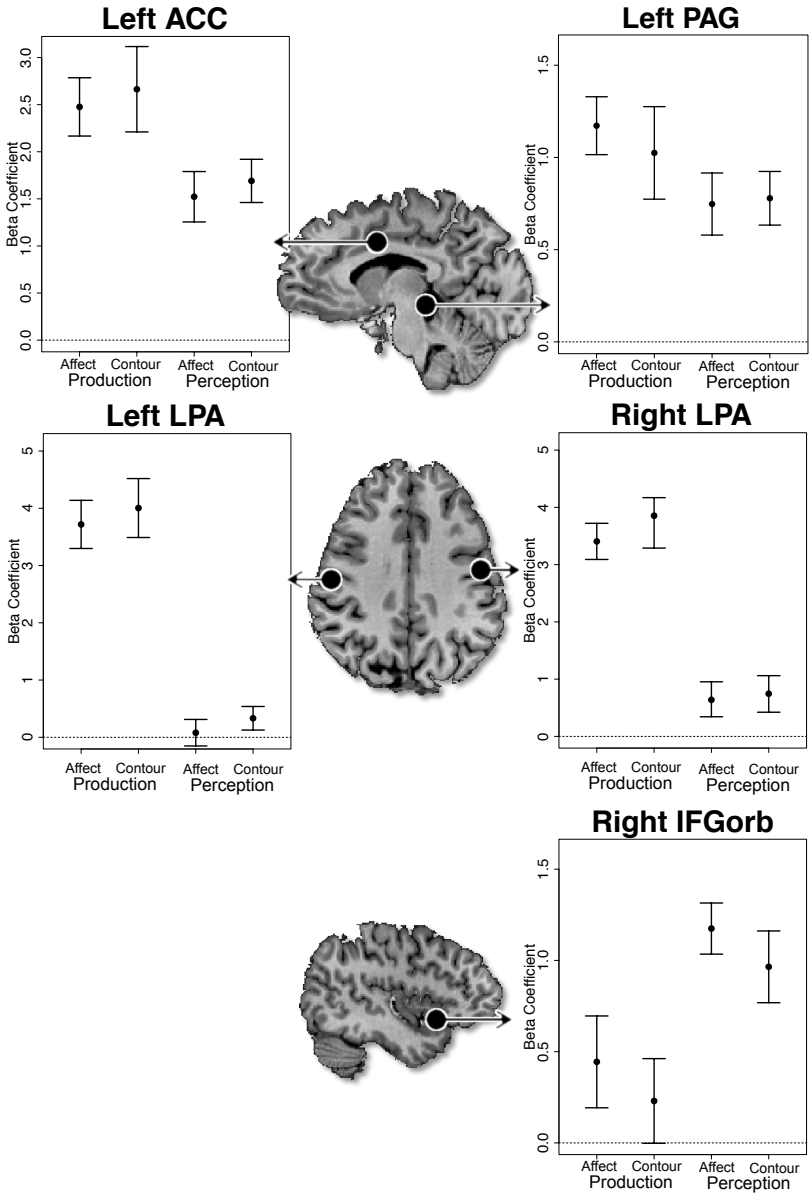

Figure 3.2: Region-of-interest analyses.

Regions of interest in the ACC, PAG, LPA, and IFG pars orbitalis are indicated by black circles on the axial and midsagittal slices. Mean beta values are plotted for each task and each ROI. Error bars mark one standard error above and below the mean. In each area, there are strong differences between the level of activation for the production and perception tasks, but no difference between the affective and pitch-contour conditions, nor are there any statistical interactions. ACC: anterior cingulate cortex; IFG: inferior frontal gyrus LPA: larynx-phonation area; PAG: periaqueductal grey.

p=0.32 and no interaction F(1,39)=0.4, p=0.54. The opposite trend was observed in the right IFGorb. A main effect of task was observed such that activation was greater for the perception than the production tasks F(1,39)=18.4, p<0.05. Neither content F(1,39)=1.5, p=0.22 nor the task-by-content interaction was significant F(1,39)=0.0002, p=0.99.

3.4.6 Functional connectivity

Activation of the ACC significantly predicted activation of the PAG − even after controlling for co-activation due to the experimental design − in both the affective t(13)=3.7, p<0.05, and contour t(13)=2.4, p<0.05 production conditions. Functional connectivity did not differ significantly between conditions t(15.8)=1.12, p=0.27. Functional connectivity was also observed between the ACC and left LPA in both tasks: affective t(13)=11.0, p<0.05, and contour t(13)=11.6, p<0.05, with no difference between tasks t(25.8)=-1.2, p<0.26. A similar trend was observed for the ACC and right LPA: affective t(13)=17.1, p<0.05, contour t(13)=12.0, p<0.05, and affect vs. contour t(21.9)=-1, p=0.31.

3.5 Discussion

The primary objective of this study was to test the presumed dissociation of affective and non-affective vocalizations as the domains of the cingulate and primary-motor vocal pathways, respectively. We had participants vocalize either affectively expressive exclamations or monosyllables expressed with

meaningless, though acoustically-matched, pitch contours. To localize brain regions involved in perceiving affective vocalizations (Belyk and Brown, 2014a; Witteman et al., 2012), we had participants listen to actor-produced affective vocalizations, and to discriminate either the valence of the emotion being expressed in the vocalization or the pitch contour.

Several analyses failed to support a functional dissociation between the two vocal-production pathways and additionally lent support to a model in which the primary process in producing affective vocalizations is the modulation of vocal parameters such as pitch: i) vocal production activated a network of brain areas that were commonly reported in studies of vocalization (Brown et al., 2009), regardless of whether the vocalizations were affective or based solely on non-affective pitch contours; ii) direct contrasts between the vocalization tasks failed to detect differences in activation anywhere in the brain; iii) ROI analyses revealed that both the cingulate vocalization pathway and LPA were equally active during both affective and non-affective vocalizing; iv) we observed functional connectivity between the ACC and PAG − the two major components of the cingulate vocalization pathway − during both vocal conditions; and v) the ACC was functionally connected with the LPA during both vocal conditions. These data suggest that the primary-motor and cingulate vocalization pathways may not be functionally dissociated, as previously suggested based non-human animal models. We further demonstrated that the IFG pars orbitalis does not participate in vocal production, contrary to our previous hypothesis (Belyk & Brown, 2014a).

3.5.1 Cingulate involvement in vocalization, affective or otherwise

While there is a paucity of research examining the production of affective vocalizations in humans, those studies that have been conducted, including the present study, are unanimous in reporting coactivation of the cingulate and LPA pathways (Aziz-Zadeh et al., 2010; Barrett et al., 2004; Wattendorf et al., 2013). While Barrett et al. (2004) observed that ACC activation correlated with vocal acoustics during expressive speech, the same was observed for the LPA. Wattendorf et al. (2013) observed that the LPA was activated during laughter, even when it was spontaneous. We further observed that the LPA and ACC are functionally connected, not only during affective but also during non-affective vocalization. While one previous study reported negative functional connectivity between these regions during simple vocal tasks (Simonyan et al., 2009), the ACC was not itself activated in those tasks, making the interpretation difficult.

While the cingulate vocalization pathway is undoubtedly essential for affective vocalization (Jürgens, 2009), evidence from the current study suggests that it is not specific for that purpose. Region-of-interest analyses revealed that the ACC and PAG were both activated and were functionally connected when participants vocalized, regardless of whether the vocalizations were affective. These findings are inconsistent with a specialization of these areas for affective vocalization, but instead suggest that they are part of a broader vocal-motor system. Interestingly, the ACC was also active during both of the perceptual judgment tasks, much like the premotor cortex, suggestive of a

role in planning rather than executing vocalizations. Previous brain imaging studies of affective vocalization in humans do strongly support the involvement of the ACC in producing affective vocalizations, but do not exclude its involvement in producing non-affective vocalizations (Barrett et al., 2004; Wattendorf et al., 2013). Indeed, the ACC is frequently activated during non-affective speech (Crosson et al., 1999; Paus et al., 1993) and singing (Brown et al., 2009). Hence, the ACC appears to participate in vocalization, affective or otherwise.

3.5.2 Pitch modulation as a common denominator

A common feature of our two vocal tasks is modulation of vocal pitch. For example, happiness and disgust are both expressed with high-pitched vocalizations, while sadness and pleasure are expressed with low-pitched vocalizations (Belyk and Brown, 2014c). Pitch level as well as pitch variability, among other vocal acoustic parameters, can further distinguish these emotions. Pitch is among the most informative acoustic cues to affective vocal expression (Banse and Sherer, 1996; Goudbeek and Scherer, 2010). The emotions expressed by vocalizations that have been manipulated to remove all but the pitch information can be discriminated above chance (Lieberman and Michaels, 1962). Sensitivity to vocal pitch cues emerges early in development. Mothers use infant-directed speech, as characterized by a raised vocal pitch (Fernald and Simon, 1984), to sooth infants, and children use pitch (among other cues) differentially so as to vary the form and intensity of their tantrums (Green et al., 2011).

Our results demonstrate that the neural system for affective vocal expression is highly similar to that for pitch production. Aziz-Zadeh et al. (2010) observed that many brain areas, including the the ACC, were active both when speakers vocally expressed an emotion and when they intoned a question, where both forms of intonation involved dynamic pitch modulation, but only one conveyed emotion. A meta-analysis of brain imaging studies of singing and simple syllable production found activation of both the LPA and cingulate vocalization area (Brown et al., 2009). Schulz et al. (2005) observed activation of the ACC and PAG, and functional connectivity between them, during non-affective speech when it was voiced, but not when it was whispered.

Taken together, the findings of the present study and the existing literature suggest that, in humans, the brain network activated by affective vocalizations may not be specific to expressing affective states, but instead may be part of the neural system for pitch modulation. This same observation applies to the perception of affective vocalizations, which overlapped the perception of pitch contours, and both of which overlapped the system for vocal production, with the exception of a few principal areas like the primary motor cortex and basal ganglia that are involved in initiating movements. This is consistent with the general property of sensorimotor systems that observing an action that one can perform activates many of the brain areas involved in executing the same action, whether it be body movements (Cross, Hamilton, and Grafton, 2006) or speech (Menenti et al., 2011). To our knowledge, only one other study has examined such a production-perception linkage in the context of prosody. Consistent with the findings of Aziz-Zadeh et al. (2010), we observed

overlapping activation throughout much the vocal-motor system during the production and perception of affective prosody.

3.5.3 The IFG pars orbitalis is specific to perception

We previously hypothesized that the IFGorb may be involved in planning the prosodic component of speech (Belyk and Brown, 2014a), in much the same way that Broca's area participates in speech planning (Watkins and Paus, 2004). This study tested that hypothesis by localizing the IFGorb with a prosodic discrimination task and tested whether it was also activated when producing affective vocalizations. Contrary to our hypothesis, ROI analyses revealed that the IFGorb was not activated by either production task and was thus specific for perception.

3.6 Conclusion

Both the cingulate vocalization pathway and the LPA were activated during vocalization, regardless of whether vocalizations expressed an emotion. This result contradicts the dominant view of these regions as separate pathways for affective and non-affective vocalization, respectively. Instead it supports a view in which the neural system for affective vocalization is very similar if not identical to the neural system for pitch production, where pitch is a principal carrier of affective signals in the voice. We further tested the hypothesis that a key brain region involved in perceiving affective prosody in human speech, the IFG pars orbitalis, also participates in producing affective prosody. The

results suggest that this region does not participate in vocal production, but instead is specific to perception.

3.7 Acknowledgments

This work was funded by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to S.B (grant no. 371336). We are grateful to Dr. Chris Aruffo for generating the vocal recordings.

3.8 References

Ackermann, H., Hage, S.R, Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences, 37*, 529−604.

Aitken, P.G. (1981). Cortical control of conditioned and spontaneous vocal behavior in rhesus monkeys. *Brain and Language, 13*(1), 171−84.

Anwander, A., Tittgemeyer, M., von Cramon, D.Y., Friederici, A.D., Knösche, T.R. (2007). Connectivity-based parcellation of Broca's area. *Cerebral Cortex, 17*(4), 816−25.

Aziz-Zadeh, L., Sheng, T., Gheytanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLOS ONE, 5*(1), e8759.

Banse, R., Sherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614−36.

Barrett, J., Pike, G.B., Paus, T. (2004). The role of the anterior cingulate cortex in pitch variation during sad affect. *European Journal of Neuroscience, 19*(2), 458−64.

Belin, P., Fillion-Bilodeau, S., Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40*(2), 531−9.

Belyk, M., Brown, S. (2014a). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience, 9*, 1395−403.

Belyk, M., Brown, S. (2014b). Somatotopy of the extrinsic laryngeal muscles in the human sensorimotor cortex. *Behavioural Brain Research, 290*, 364−71.

Belyk, M., Brown, S. (2014c). The acoustic correlates of valence depend on emotion family. *Journal of Voice, 28*(4), 523.e9−e18.

Brown, S., Laird, A.R., Pfordresher, P.Q., Thelen, S.M., Turkeltaub, P., Liotti, M. (2009). The somatotopy of speech: Phonation and articulation in the human motor cortex. *Brain and Cognition, 70*(1), 31−41.

Brown, S., Ngan, E., Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex, 18*(4), 837−45.

Cross, E.S., Hamilton, A.F.D.C., Grafton, S.T. (2006). Building a motor simulation de novo: observation of dance by dancers. *Neuroimage, 31*(3), 1257−67.

Crosson, B., Sadek, J.R., Bobholz, J.A., Gökay, D., Mohr, C.M., Leonard, C.M., et al. (1999). Activity in the paracingulate and cingulate sulci during word generation: An fMRI study of functional anatomy. *Cerebral Cortex, 9*(4), 307−16.

Fairbanks, G., Pronovost, W. (1938). Vocal pitch during simulated emotion. *Science*, 88(2286), 382−83.

Farag, T., Andics, A., Devecseri, V., Kis, A., Gcsi, M., dm, M. (2014). Humans rely on the same rules to assess emotional valence and intensity in conspecifics and dog vocalizations. *Biology Letters, 10*, 1−5.

Fernald, A., Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104−13.

Goudbeek, M., Scherer, K. (2010). Beyond arousal: Valence and potency control cues in the vocal expression of emotion. *Journal of the Acoustical Society of America, 128*(3), 1322−36.

Green, J.A., Whitney, P.G., Potegal, M. (2011). Screaming, yelling, whining, and crying: Categorical and intensity differences in vocal expressions of anger and sadness in children's tantrums. *Emotion, 11*(5), 1124−33.

Hast, M.H., Fischer, J.M., Wetzel, A.B. (1974). Cortical motor representation of the laryngeal muscles in macaca mulatta. *Brain, 73*, 229−40.

Iwatsubo, T., Kuzuhara, S., Kanemitsu, A. (1990). Corticofugal projections to the motor nuclei of the brainstem and spinal cord in humans. *Neurology, 40*, 309−12.

Jürgens, U. (1974). On the elicitability of vocalization from the cortical larynx area. *Brain Research, 81*(3), 564−6.

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews, 26*(2), 235−58.

Jürgens, U. (2009). The neural control of vocalization in mammals: A review. *Journal of Voice, 23*(1), 1−10.

Jürgens, U., Ehrenreich, L. (2007). The descending motorcortical pathway to the laryngeal motoneurons in the squirrel monkey. *Brain Research, 1148*, 90−5.

Jürgens, U., Kirzinger, A., von Cramon, D. (1982). The effects of deep-reaching lesions in the cortical face area on phonation: A combined case report and experimental monkey study. *Cortex, 18*, 125−39.

Jürgens, U., Ploog, D. (1970). Cerebral representation of vocalization in the squirrel monkey. *Experimental Brain Research, 554*, 532−54.

Jürgens, U., Pratt, R. (1979a). Role of the periaqueductal grey in vocal expression of emotion. *Brain Research, 167*, 367−78.

Jürgens, U., Pratt, R. (1979b). The cingular vocalization pathway in the squirrel monkey. *Experimental Brain Research, 510*, 499−510.

Kirzinger, A., Jürgens, U. (1982). Cortical lesion effects and vocalization in the squirrel monkey. *Brain Research, 233*, 299−15.

Kuypers, H.G.J.M. (1958). Corticobulbar connexions to the pons and lower brain-stem in man. *Brain, 81*(3), 364−88.

Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Chui, W., et al. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology, 4*, 1−8.

Lieberman, P., Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *Journal of the Acoustical Society of America, 34*(7), 922−7.

Loucks, T.M.J., Poletto, C.J., Simonyan, K., Reynolds, C.L., Ludlow, C.L. (2007). Human brain activation during phonation and exhalation: Common volitional control for two upper airway functions. *Neuroimage, 36*(1), 131−43.

Menenti, L., Gierhan, S. M.E., Segaert, K., Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychological Science, 22*(9), 1173−82.

Myers, R.E. (1976). Comparative neurology of vocalization and speech: Proof of a dichotomy. *Annals of the New York Academy of Sciences, 280*(1), 745−57.

Owren, M.J., Amoss, R.T., Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology, 73*(6), 530−44.

Paus, T., Petrides, M., Evans, A.C., Meyer, E. (1993). Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses: A positron emission tomography study. *Journal of Neurophysiology, 70*(2), 453−69.

Penfield, W., Boldrey, E. (1937). Somatic motor and sensory representations in the cerebral cortex of man as studied by electrical stimulation. *Brain, 60*, 389−443.

Petkov, C.I., Jarvis, E.D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience, 4*, 1−24.

Price, J.L. (1999). Prefrontal cortical networks related to visceral function and mood. *Annals of the New York Academy of Sciences, 877*, 383−96.

R Core Development Team (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Sauter, D.A., Eisner, F., Ekman, P., Scott, S.K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences, 107*(6), 2408−12.

Scheiner, E., Hammerschmidt, K., Jürgens, U., Zwirner, P. (2004). The influence of hearing impairment on preverbal emotional vocalizations of infants. *Folia Phoniatrica et Logopaedica, 56*(1), 27−40.

Scheiner, E., Hammerschmidt, K., Jürgens, U., Zwirner, P. (2006). Vocal expression of emotions in normally hearing and hearing-impaired infants. *Journal of Voice, 20*(4), 585−604.

Scherer, K.R., Wallbott, H.G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology, 66*(2), 310−28.

Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication, 40*, 99−116.

Schulz, G.M., Varga, M., Jeffires, K., Ludlow, C.L., Braun, A.R. (2005). Functional neuroanatomy of human vocalization: An H215O PET study. *Cerebral Cortex, 15*(12), 1835−47.

Simon-Thomas, E.R., Keltner, D.J., Sauter, D., Sinicropi-Yao, L., Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion, 9*(6), 838−46.

Simonyan, K., Jürgens, U. (2003). Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Research, 974*(1-2), 43−59.

Simonyan, K., Ostuni, J., Ludlow, C.L., Horwitz, B. (2009). Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *Journal of Neuroscience, 29*(47), 14912−23.

Turken, A.U., Dronkers, N.F. (2011). The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses. *Frontiers in Systems Neuroscience, 5*, 1−20.

Sutton, D., Larson, C., Lindeman, R. C. (1974). Neocortical and limbic lesion effects on primate phonation. *Brain Research, 71*(1), 61−75.

Sutton, D., Trachy, R.E., Lindeman, R.C. (1981). Primate phonation: Unilateral and bilateral cingulate lesion effects. *Behavioural Brain Research, 3*(1), 99−114.

Watkins, K., Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience, 16*(6), 978−87.

Wattendorf, E., Westermann, B., Fiedler, K., Kaza, E., Lotze, M., Celio, M.R. (2013). Exploration of the neural correlates of ticklish laughter by functional magnetic resonance imaging. *Cerebral Cortex, 23*, 1280−9.

Witteman, J., Van Heuven, V.J.P., Schiller, N.O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia, 50*(12), 2752−63.

Yeterian, E. H., Pandya, D. N., Tomaiuolo, F., Petrides, M. (2012). The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex, 48*(1), 58−81.

Chapter 4

The neural basis of vocal pitch imitation in humans

Michel Belyk, Peter Q. Pfordresher, Mario Liotti & Steven Brown.

Submitted to the Journal of Cognitive Neuroscience.

4.1 Abstract

Vocal imitation is a phenotype that is unique to humans among all primate species, and so an understanding of its neural basis is critical to explaining the emergence of both speech and song in human evolution. Two principal neural models of vocal imitation have emerged from a consideration of non-human animals. In monkeys, mirror neurons in the homologous region of Broca's area are important for gestural imitation, leading to the hypothesis that this region

may have been coopted to support vocal imitation in humans. An alternative hypothesis derived from the study of songbirds suggests that the cortico-striate motor pathway performs processes critical for vocal imitation. Using functional magnetic resonance imaging with a sparse event-related sampling design, we investigated the neural basis of vocal imitation in humans by comparing imitative vocal production of pitch sequences with both non-imitative vocal production and pitch discrimination. The strongest difference between these tasks was found in the putamen bilaterally, providing a striking parallel to the role of the analogous region in songbirds. Other areas preferentially activated during imitation included the larynx motor cortex, subcentral gyrus, and supplementary motor area, which together outline the cortico-striate motor loop. No differences were seen in Broca's area. The cortico-striate system thus appears to be the central pathway for vocal imitation in humans, as predicted from an analogy with songbirds.

4.2 Introduction

Although most vertebrates have the capacity to vocalize, very few species have the ability to learn their vocal repertoires through imitation the way that humans so effortlessly do. Among the principal exceptions are dolphins (King & Sayigh, 2013), whales (Noad, Cato, Bryden, Jenner, & Jenner, 2000), and bats (Knörnschild, Nagy, Metz, Mayer, & von Helversen, 2010). Limited evidence also suggests that elephants (Poole, Tyack, Stoeger-Horwath, & Watwood, 2005; Stoeger et al., 2012), seals (Ralls, Fiorelli, & Gish, 1985; Sanvito, Galimberti, & Miller, 2007), and mice (Arriaga & Jarvis, 2013) may be capable

of vocal imitation, although the evidence remains sparse. Vocal imitation in humans is important not only during childhood development for the establishment of large and flexible acoustic repertoires for speech and music (Kuhl & Meltzoff, 1996; Papousek, 1996; Poulson, Kymissis, Reeve, Andreators, & Reeve, 1991; Studdert-Kennedy, 2000; Trehub, 2001), but also throughout adult life for the ability to, for example, learn musical melodies and produce the sounds of a foreign language.

While theories of vocal imitation are diverse, they tend to agree on a core set of processes related to the sensorimotor translation of perceived sounds (Berkowska & Dalla Bella, 2009; Dalla Bella & Berkowska, 2009; Hutchins & Moreno, 2013; Pfordresher et al., 2015; Pfordresher & Mantell, 2014). As shown in Figure 1, vocal imitation requires that an individual perceive a target sound, map the acoustic properties of the target onto phonatory and articulatory motor commands through a process of causal inverse modeling, and then execute those commands to vocally reproduce the target sound.

There is a widespread population of individuals − colloquially known as "tone deaf" individuals, but more accurately described as "poor-pitch singers" − who have a specific deficit in the sensorimotor translation involved in vocal pitch imitation. Poor-pitch singers are often accurate at encoding auditory stimuli − as demonstrated by performance on pitch-discrimination tasks (Pfordresher & Brown, 2007) − but are deficient in translating that internal model into an appropriate motor signal so as to match the acoustic properties of the model (Pfordresher & Mantell, 2014). Their deficit is neither sensory nor motor, but rather sensorimotor (i.e., imitative). This is suggestive of a specific

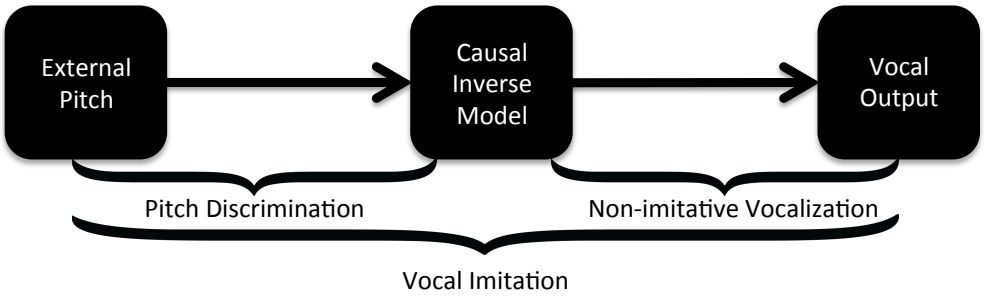

Figure 4.1: Model of vocal pitch imitation.
In vocal imitation, an external pitch stimulus is perceived, converted to a motor code via a causal inverse model, and this motor program is then executed at the level of the larynx.

deficit in mapping auditory percepts onto phonatory motor commands.

The standard neural model of vocal imitation in humans is based on the neurological literature involving aphasic patients. The classic Wernicke-Geschwind model posits that auditory information is transmitted from the posterior part of the superior temporal gyrus (pSTG) to the inferior frontal gyrus (IFG) via the arcuate fasciculus (AF), and then presumably to the motor cortex for vocal execution, although the model does not specify this final step. Lesions to the AF, which effectively disconnect the pSTG from the IFG, cause deficits specific to vocal imitation, with spared speech comprehension and otherwise fluent speech production. These observations led to the hypothesis that Broca's area may convert auditory information from the temporal lobe into articulatory-motor commands, which are then executed by the primary motor

cortex (Geschwind, 1970). While the Wernicke-Geschwind model only describes this pathway in the left hemisphere in the context of language processing, the homologous pathway in the right hemisphere is structurally similar, and vocalization often engages a bilateral audio-vocal network, whether in the context of producing speech or non-speech sounds (Brown, Ngan, & Liotti, 2008; Chang, Kenney, Loucks, Poletto, & Ludlow, 2009; Mashal, Solodkin, Dick, Elinor Chen, & Small, 2012).

Vocal imitation is first and foremost a process of sensorimotor translation of an acoustic stimulus. Several neural models of vocal imitation have taken their lead from theories of gestural imitation based on mirror neurons. Mirror neurons are cells that have been described in the brains of monkeys that fire both when an animal perceives and produces a particular action (di Pallegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992), and have been proposed to play a key role in gestural imitation. While the single-cell recording studies necessary to demonstrate the existence of mirror neurons in the human brain have not been conducted, neuroimaging studies have identified brain areas that constitute populations of cells that together display mirror-like properties (Gazzola & Keysers, 2009). Among these putative mirror-neuron regions is the posterior portion of Broca's area, consisting of Brodmann Area (BA) 44 in the IFG pars opercularis. This region is activated both when viewing manual gestures and when producing them from memory (Iacoboni et al., 1999). However, activation is greatest when imitating novel gestures, suggesting a specific role for this area in gestural imitation. Although a meta-analysis of the gestural imitation literature questioned the reliability of such an imitation effect in the

IFG pars opercularis (Molenberghs, Cunnington, & Mattingley, 2009), repetitive transcranial magnetic stimulation of this region disrupts manual imitation (Heiser, Iacoboni, Maeda, Marcus, & Mazziotta, 2003). Such findings, combined with the well known role of Broca's area in speech planning (Papoutsi et al., 2009), have led researchers to speculate that Broca's area may also be a key region for vocal learning via imitation (Iacoboni et al., 1999; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996).

Certain species of birds that possess the capacity for vocal production learning provide an alternative neural model for vocal imitation. In contrast to monkeys, three lineages of birds, namely parrots, hummingbirds and songbirds, are capable of learning novel vocalizations through vocal imitation (Nottebohm, 1972). The vocal system of vocally-imitating birds, particularly songbirds, has been studied extensively (Jarvis, Güntürkün, & Bruce, 2005). The avian song system consists of two pathways: a descending vocal-motor pathway and a forebrain-striatal loop. While lesions to the descending vocal-motor pathway profoundly disrupt song production (Nottebohm, Stokes, & Leonard, 1976), lesions to the forebrain-striatal loop disrupt vocal imitation and song learning, but spare the production of songs that have already been learned (Bottjer, Miesner, & Arnold, 1984; Sohrabji, Nordeen, & Nordeen, 1990). Neurophysiological evidence suggests that neurons along the forebrain-striatal loop compute causal inverse models that map target sounds onto the motor commands that reproduce them (Giret, Kornfeld, Ganguli, & Hahnloser, 2014). The brain areas that comprise the two songbird vocal pathways have analogues in the human brain (see Jarvis, Güntürkün, & Bruce, 2005

for a review), and these analogues are also active when humans sing (Brown, Martinez, Hodges, Fox, & Parsons, 2004). Indeed Area X, a key node in the songbird forebrain-striatal loop, shares molecular specializations with the human putamen (Pfenning et al., 2014). While the basal ganglia as a whole are highly conserved across vertebrates, species may develop novel modules as they evolve new behaviours (Grillner, Robertson, & Stephenson-Jones, 2013). Humans and songbirds may have convergently evolved novel modules in the basal ganglia that support vocal imitation.

Vocal imitation of pitch is an ideal medium for examining audiovocal matching since pitch is a highly salient component of vocal communication that can be measured with greater simplicity and precision than either gestural or articulatory imitation. The first human neuroimaging study on vocal pitch imitation was that of Brown et al. (2004), although that study lacked the experimental controls to make specific claims about vocal imitation, as compared to vocalization per se. The present study attempted to compare imitative vocalization with the highly matched conditions of non-imitative vocalization and pitch discrimination, using sparse temporal sampling (Hall et al., 1999) so as to measure behavioural performance in the scanner. The principal aim was to shed light on the unique ability of humans among primates to perform vocal imitation by comparing the two competing hypotheses that either Broca's area or the cortico-striate system supports vocal imitation in humans, as predicted by the "gestural imitation" and "avian song-system" animal models, respectively. In the imitation condition, subjects listened to novel melodies and then imitated them vocally, thereby engaging all of the

processes shown in the Figure 1. In a non-imitative vocalization condition, participants were visually cued to sing highly familiar melodies, thereby engaging pre-existing motor commands. Finally, in a pitch-discrimination condition, participants heard pitch sequences and had to detect pitch changes, thereby engaging auditory but not vocal-motor processes.

4.3 Methods

4.3.1 Participants

Fourteen participants (median age 25, range 19-48, 7 female, 1 left handed) were recruited at Simon Fraser University. Participants were prescreened to verify that they were accurate vocal imitators, although with stimuli modified to resemble those used in the present study (see below). All participants had absolute note errors of less than one semitone (i.e., 100 cents), on average, which was the criterion for accurate imitation established in Pfordresher & Brown (2007). One participant was excluded due to undiagnosed hydrocephalus.

4.3.2 Stimuli and procedure

Participants completed each of three tasks twice in separate runs in random order. For each task, the same stimuli were presented across runs, but in counterbalanced pseudorandom order. Each experimental task consisted of a visual cue, a four-note auditory stimulus, a response period, and a variable delay before image acquisition (Figure 2). Experimental trials alternated with

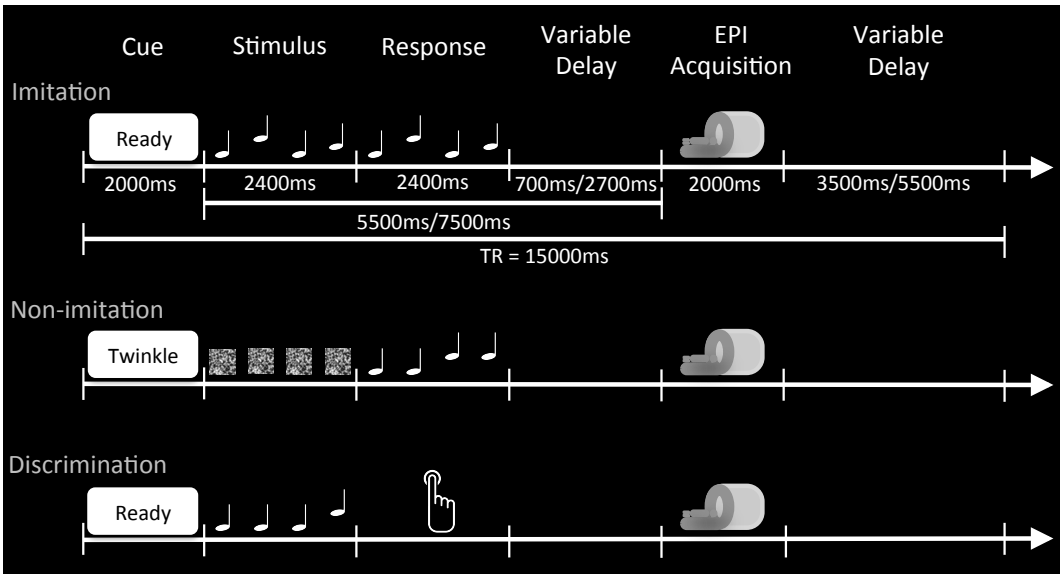

Figure 4.2: Trial timing.

The timing of trials within each of the three conditions is depicted. In the vocal imitation condition, participants heard novel four-note melodies, and then imitated them vocally. In the non-imitative vocalization condition, participants were visually cued with the name of a highly familiar melody, heard four task-irrelevant white-noise bursts, and then sang the first four notes of the target melody. In the pitch-discrimination condition, participants heard a series of three identical notes followed by a fourth note, and then indicated on a response pad whether the fourth note was the same or different than the preceding three. Based on the use of a sparse temporal sampling design, EPI images were collected after each trial. Hence, participants performed all tasks in the absence of scanner noise.

a rest condition, during which participants fixated on a crosshair. The eyes were kept open in all scans.

Vocal imitation task

Eighteen novel four-note melodies were synthesized in a vocal timbre on the vowel /u/ using Vocaloid (Leon, Zero-G Limited, Okehampton, U.K.). All melodies were isochronous with 600 ms inter-onset intervals, with a 50 ms 10

dB fade-in and drop-off. Notes ranged from A2 (110Hz) to E3 (164.81Hz) for males and from A3 (220Hz) to E4 (329.63Hz) for females. Stimuli were generated in equal numbers with three levels of complexity, in accordance with the stimuli of Pfordresher & Brown (2007). "Note" stimuli consisted of a sequence of four identical notes. "Interval" stimuli consisted of two doublets of notes with a single interval between the first and second doublet (e.g., AAEE). "Melody" stimuli consisted of a series of non-repeating notes (e.g., ABC#E). A "Ready" screen was displayed 2s before the onset of a stimulus in order to indicate that a trial was about to begin. The target melody was presented for 2400 ms followed by a 2400 ms response period, during which participants were instructed to imitate the target melody.

Non-imitative vocalization task

Participants were visually cued with the name of a familiar melody and instructed to vocalize the first four notes of the melody. Participants vocalized either a monotone sequence (i.e., four identical pitches), "Twinkle, Twinkle", or "Mary Had a Little Lamb". These stimuli matched the number of note-changes in the note, interval, and melody stimuli, respectively, of the vocal imitation task. After the verbal cue, four white-noise bursts were presented that matched the amplitude and duration of the stimulus melodies of the vocal imitation task. This was done to match the level of auditory stimulation that was present in the vocal-imitation and pitch-discrimination conditions. Participants were instructed to produce the familiar melodies from memory in a comfortable part of their vocal range after the white noise bursts were

completed.

Pitch discrimination task

Eighteen four-note melodies were synthesized in the same manner as the target melodies of the vocal imitation task. The first three notes of each melody were A2 for males or A3 for females. On half of the trials, the final note was identical to the initial notes. In the remaining trials, the final note was 25, 50, 100, 200, 400 or 600 cents higher or lower than the initial notes (where 100 cents = 1 equal-tempered semitone). Participants pressed a button to indicate whether the final note was identical or not to the initial notes. Button presses were recorded on an MRI-compatible button box with the index and middle fingers of the right hand, where the "same" and "different" options were counterbalanced across subjects.

4.3.3 Imitation analysis

Sung melodies were recorded from participants in the scanner using an MRI-compatible microphone that fed into the Avotek patient-communication system, itself connected to a laptop computer running Adobe Audition. Sung melodies from the scanner were then subjected to acoustic analysis. The pitch of each sung note was extracted using the autocorrelation algorithm as implemented in Praat (Boersma & Weenink, 2011) and compared to the corresponding notes of each target melody. The intervals of the target and sung melodies were calculated as the difference between adjacent notes in the target and sung melodies, respectively. Performance on the vocal imitation task in

the scanner was assessed by both the accuracy and precision of both the notes and the melodic intervals, as described in (Pfordresher et al., 2010). Accuracy was measured as the mean signed difference between the notes or intervals of sung melodies and those of the target melodies. Precision was measured as the standard deviation of note and interval errors across pitch classes.

4.3.4 Magnetic resonance imaging

Magnetic resonance images were acquired with a Phillips 3-Tesla MRI. Functional images sensitive the blood oxygen level-dependent (BOLD) signal were collected with gradient echo sequences according to a sparse event-related sampling design (Hall et al., 1999). Samples were collected 5500 or 7500 ms after stimulus onset on alternating trials to eliminate scanner noise during auditory-stimulus presentation and vocalization, as well as to minimize movement-related artifacts during image acquisition. The eyes were kept open during the scans. Imaging parameters were: repetition time = 15000 ms, acquisition time = 2000 ms, echo time = 33 ms, flip angle = 90 degrees, 36 slices, slice thickness = 3 mm, gap = 1 mm, in-plane resolution 1.875 mm by 1.875 mm, matrix = 128 x 128, and field of view = 240 mm. A total of 39 whole-brain volumes were collected per scan. The first three were discarded, leaving 36 volumes, corresponding to 18 alternations between task and rest trials. A T1-weighted image with 1 mm isotropic voxels and field of view 256 mm by 256 mm by 170 mm was also collected for image registration.

4.3.5 Image analysis

Functional scans were analyzed with Brain Voyager QX 2.8 supplemented with NeuroElf (neuroelf.net). Each functional scan was spatially smoothed with a Gaussian kernel of 4 mm full-width-half-maximum, and high-pass filtered with a cut-off frequency of 0.0078125 Hz (or 1/128s). Each sample was realigned with the first sample to correct for head motion.

To localize the basic audio-vocal network, we performed a three-way conjunction between vocal imitation, non-imitative vocalization, and pitch discrimination. To further identify vocal-motor-related activations, we performed a conjunction of the contrasts [Imitation > Discrimination] ∩ [Non-imitation > Discrimination]. These contrasts were corrected for multiple comparisons with a False Discovery Rate (FDR) of $q<0.01$ and an additional cluster threshold of $k>12$.

To identify regions of the vocal network that were preferentially activated by vocal imitation, we performed a conjunction of the contrasts [Imitation>Non-imitation] ∩ [Imitation>Discrimination]. This conjunction identified brain regions that were more active during vocal imitation than both non-imitative vocalization and pitch discrimination. A cluster threshold of $k>18$ was applied to an uncorrected p-threshold of 0.05 in order to preserve a cluster-level family-wise error rate of $p<0.05$, as determined by Monte Carlo simulation.

Region-of-interest analysis

We identified functionally-localized regions-of-interest (ROIs) based on 5mm cubes drawn around the activation peaks of each brain region identified in the

vocal-imitation conjunction analysis. Beta coefficients from first-level analyses for all participants were extracted from each brain area for each condition and compared using the general linear model as implemented in R (R Core Development Team, 2014).

4.4 Results

4.4.1 Behavioural data

The mean accuracy score of vocal-imitation performance in the scanner, combined across note and interval measurements, was 44.5 (SD 17.0) cents. The mean precision of imitation was 66.4 (SD 41.6) cents. This suggests that the subjects were accurate and precise imitators, according to established criteria for these parameters (Pfordresher et al., 2010; Pfordresher & Brown, 2007). These measurements replicated imitation performance during the prescreening experiments. Median performance on the pitch discrimination task was 94.4%.

4.4.2 Imaging data

Vocal imitation, non-imitative vocalization, and pitch discrimination all activated a basic audio-vocal network. A conjunction between these three conditions (Figure 3) revealed shared activations in bilateral Heschl's gyrus (BA 41) extending into the pSTG (BA 42 and 22), orofacial premotor cortex (BA 6), IFG pars opercularis (BA 44), anterior insula (BA 13), putamen, thalamus, and lateral cerebellum. Shared activations were observed in the bilateral

SMA, ACC, and cerebellar vermis. The shared audio-vocal areas identified in this conjunction reflect a neural system for the internal encoding of melodies resulting from either online perception or access from long-term stores.

A conjunction of the contrasts [Imitation>Discrimination] ∩ [Non-imitation >Discrimination] revealed a set of regions preferentially activated during vocal production. This extended the abovementioned network to include the bilateral larynx-phonation area and subcentral gyrus, as well as bilateral Heschl's gyrus (BA 41) and right SMA.

Vocal imitation

The conjunction [Imitation>Non-imitation] ∩ [Imitation>Discrimination] revealed a subset of the audio-vocal network that was more active during vocal imitation than both non-imitative vocalization and pitch discrimination (Figure 4). These areas included the right larynx-phonation area (BA 4/3), left sub-central gyrus (BA 6/43), the bilateral SMA (BA 6), and bilateral putamen. Notably, Broca's area was not among the areas revealed by this analysis. All of these areas were also present in each condition individually (as seen in figure 3), suggesting that, while they were preferentially engaged by vocal imitation, they were by no means specific to that task. ROI analyses (Figure 5) of these regions indicated that they were activated in all three tasks − not just the imitation task − suggestive of a potential species difference from the songbird.

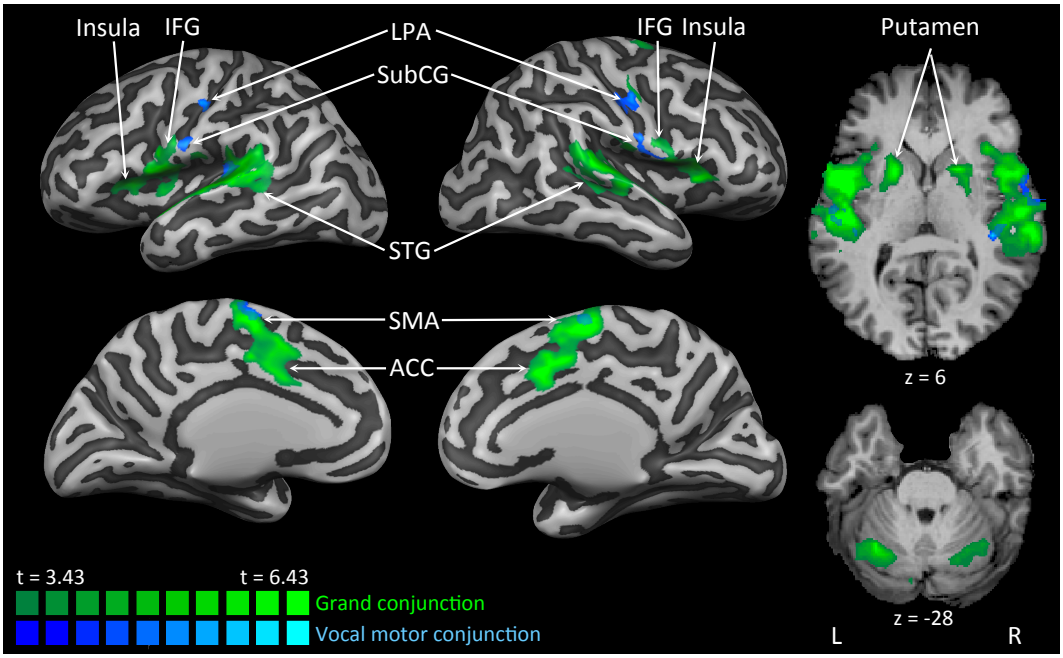

Figure 4.3: The audio-vocal network.

Activation maps for the conjunction of vocal imitation, non-imitative vocalization and pitch discrimination (green) show those elements of the audio-vocal system that are activated during all three tasks. The conjunction of contrasts [Imitation > Discrimination] ∩ [Non-imitation > Discrimination] (blue) shows brain areas that were preferentially engaged during vocalization. Both maps are threshold at FDR<0.01 k>12. Abbreviations: ACC: anterior cingulate cortex; IFG: inferior frontal gyrus; LPA: larynx-phonation area; SMA: supplementary motor area; SubCG: subcentral gyrus.

Table 4.1: The audio-vocal network.

Location of peak voxels for the three experimental contrasts against fixation. After each anatomical name in the brain region column, the Brodmann number for that region is listed. The columns labeled as x, y, and z contain the Talairach coordinates for the peak of each cluster reaching significance at FDR p>0.005 with cluster threshold k>24. Abbreviations: ACC: anterior cingulate cortex; IFG: inferior frontal gyrus; pSTG: superior temporal gyrus; SMA: supplementary motor area.

	Imitation					Non-imitation					Discrimination				
Brain Regions	x	y	z	Voxels	t-value	x	y	z	Voxels	t-value	x	y	z	Voxels	t-value
Frontal Lobe															
SMA (BA6)	3	-7	59	422	36.0	3	-7	59	261	19.9	7	-7	56	35	11.6
Ant. CingG (BA 32)						3	11	37	71	12.8	3	10	42	483	13.6
Pre/PostCG (BA 6/4/3)	-54	-6	23	45	17.7	57	-11	40	45	9.8	-46	-5	18	40	14.0
	-44	-18	41	51	9.8	-50	-14	23	30	11.8	-49	-30	37	45	10.6
	53	-19	42	98	10.6	-44	-15	46	50	7.8	-53	-22	17	349	13.2
	3	-45	66	32	7.0						-58	-25	29	56	12.8
ant. Ins. (BA 13)	-37	17	19	25	9.5						-40	17	16	44	10.1
IFG (BA 44)	-52	1	12	190	23.6	-49	1	12	195	21.3	36	16	14	239	20.4
	57	-1	10	523	24.6						-52	4	12	260	15.3
MFG											44	34	33	27	8.3
Temporal Lobe															
A1 (BA 41)	-48	-22	13	632	29.5						-50	-33	17	73	11.7
						45	-24	8	27	19.5					
STG (BA 42)	-42	-39	19	38	25.7	-58	-25	15	485	29.8					
	59	-28	12	236	23.5	59	-19	10	558	26.0	59	-28	19	263	13.1
STG (BA 22)						51	-10	9	35	17.9					
Parietal Lobe															
IPL (BA 40)						0	-51	65	48	8.3	-38	-48	54	50	10.2
											-41	-49	47	119	10.0
											48	-47	33	115	10.0
Post CingG (BA 23)											-6	-23	26	101	9.8
Subcortical															
Striatum	-23	10	16	214	15.3	-16	-5	6	78	9.9	-23	3	13	225	12.4
						-19	7	13	28	8.0					
						-19	-2	22	29	7.6	22	-1	15	83	10.6
	16	-1	2	195	13.0	16	-4	6	40	10.0	31	-5	9	41	11.2
Thalamus	15	-25	1	40	11.3	15	-25	1	40	11.3	-19	-16	15	25	7.4
Cerebellum	0	-75	-20	52	11.8	0	-78	-17	110	12.6	-3	-53	-5	51	8.7
	-3	-63	-5	53	11.4	-28	-57	-17	72	10.2	-39	-48	-23	66	8.6
	-27	-54	-17	29	8.5	21	-59	-19	95	10.7	33	-51	-21	40	7.3
	36	-48	-23	53	9.7	44	-59	-24	132	10.8					
	12	-59	-13	54	9.6										

Table 4.2: Vocal imitation.

Location of peak voxels for the conjunction of high-level contrasts [Imitative vocalization > Non-imitative vocalization] ∩ [Imitation > Discrimination]. After each anatomical name in the brain region column, the Brodmann number for that region is listed. The columns labeled as x, y, and z contain the Talairach coordinates for the peak of each cluster reaching significance at a threshold of $p<0.05$ and $k<18$, as selected by Monte Carlo simulation. Abbreviation: SMA: supplementary motor area.

	Conjunction of Contrasts				
Brain Regions	x	y	z	Voxels	t-value
Frontal Lobe					
Larynx-phonation area (BA 4/3)	53	-14	36	291	3.78
Sub-central gyrus (BA 6/43)	-49	1	6	231	4.27
SMA (BA 6)	-1	-8	63	268	3.18
Subcortical					
Striatum	11	13	0	140	3.13
Striatum	-22	-5	18	358	4.08

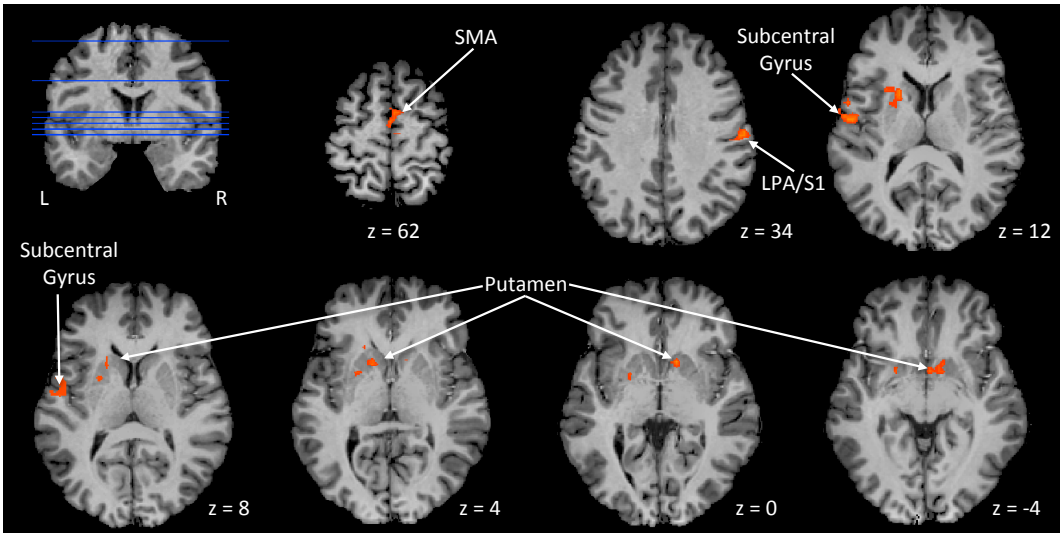

Figure 4.4: Vocal imitation.

A whole brain map display of the conjunction of high-level contrasts [Imitative vocalization > Non-imitative vocalization] ∩ [Imitation > Discrimination] depicting areas of the brain activated during vocal imitation with a threshold of $p<0.05$ and $k<18$, as selected by Monte Carlo simulation. Blue lines on the coronal slice ($y=0$) indicate the levels at which axial slices were taken. Abbreviations: LPA: larynx-phonation area; SMA: supplementary motor area; S1: primary somatosensory cortex.

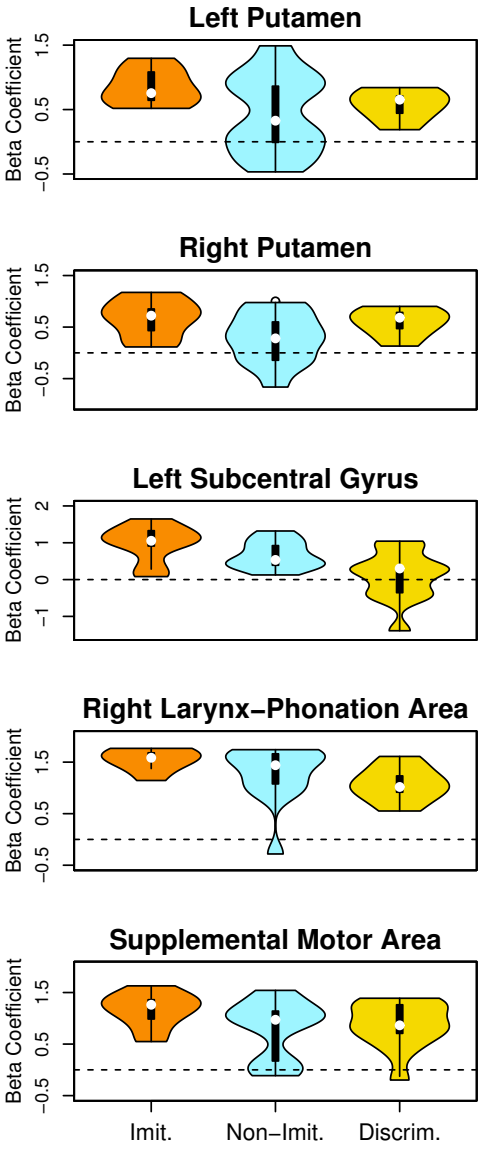

Figure 4.5: Descriptive plots for region of interest analyses.

Violin plots show the distribution of beta coefficients for imitative vocalization, non-imitative vocalization, and pitch discrimination in each brain area that was preferentially engaged by vocal imitation. The dashed horizontal line marks beta values of zero in each plot. These plots demonstrate that, while vocal imitation preferentially engaged these regions, they were not specific to imitation. This suggests that this cortico-striate system contributes to both the encoding and production phases of vocal imitation, in addition to any imitation-specific processes.

4.5 Discussion

In order to shed light on the unique ability of humans among primates to perform vocal imitation, we conducted a targeted comparison between imitative vocalization and the closely matched tasks of non-imitative vocalization and auditory-discrimination so as to identify brain areas preferentially activated by imitation. We did so using accurate imitators with a sparse temporal sampling fMRI protocol that both created a silent environment for the participants to perform the task and that permitted us to record vocal behaviour in the scanner. The results failed to show a significant imitative effect in Broca's area but instead demonstrated a clear effect in the cortico-striate pathway, including the putamen, SMA, and larynx motor cortex, suggesting that these regions are preferentially engaged during vocal imitation.

These results are consistent with an extensive literature showing that the basal ganglia function in the acquisition of novel motor sequences (Shmeulof and Krakaeur, 2011). Importantly, ROI analyses showed that the putamen was activated both when perceiving pitches and when singing them, hence creating an important link between these two phases of vocal imitation. Consistent with previous research (Brown & Martinez, 2007), all three tasks, including the non-vocal pitch-discrimination task, activated an overlapping set of brain regions that contained the majority of areas comprising the audio-vocal network. Only the larynx motor areas of the primary motor cortex and subcentral gyrus were specifically activated during vocal production.

The classical model of vocal imitation in humans, namely the Wernicke-Geschwind model (Geschwind, 1970), implicates Broca's area as a key node in

the imitative pathway. According to this model, the arcuate fasciculus relays auditory information from the temporal lobe to speech-planning areas in the frontal lobe. Indeed, lesions to the AF, usually accompanied by temporal-lobe grey matter damage (Buchsbaum et al., 2011; Damasio & Damasio, 1980), can cause conduction aphasia. This disorder is characterized by imitation-specific speech deficits, with sparing of both the production and comprehension of speech. However, we observed no specificity for vocal imitation in the brain areas that lie at either end of the AF (i.e., the pSTG and IFG). These findings suggest that, while the AF pathway may be necessary for relaying auditory information to the motor system, critical processes specific to vocal imitation occur downstream of this pathway.

We suggest that one such process is the computation of causal inverse models in the basal ganglia. Stronger activations for imitation compared to non-imitative production were found in several regions of the vocal motor network. Most notably, the putamen, which is analogous to songbird Area X − itself a key node in the vocal-imitation pathway of songbirds − was more active during vocal imitation than either non-imitative vocalization or pitch discrimination, although both of these latter tasks also activated the putamen to some degree. This imitation effect is consistent with neurophysiological work in the songbird showing that Area X receives afferents from pallial mirror neurons (Prather, Peters, Nowicki, & Mooney, 2008) and is a strong candidate for being the source of the causal inverse models that relate target sounds to motor commands (Giret, Kornfeld, Ganguli, & Hahnloser, 2014).

While the current study focused on the imitation of vocal pitch, there is

a literature devoted to speech repetition, particularly of pseudo-words, that may engage some of the processes involved in vocal imitation. However, few studies of pseudo-word repetition have been designed to address imitation directly. Nonetheless, the results of these studies are broadly consistent with our finding that the putamen is preferentially engaged during vocal imitation. For example, studies of pseudo-word processing have shown that the putamen is activated by pseudo-word repetition (Peeva et al., 2010), and that the level of activation decreases with practice (Rauschecker, Pringle, & Watkins, 2008), consistent with a transition from motor learning to motor-program retrieval. Separate subdivisions of the putamen may underlie imitating novel pseudo-words compared to retrieving motor commands to produce well-known real words (Hope et al., 2014). One previous fMRI study similarly compared repeating native-language pseudo-words to repeating foreign words that contained unfamiliar phonemes (Simmonds, Leech, Iverson, & Wise, 2014). While pseudo-words contain novel sequences of phonemes, foreign words contain novel phonemes, which are themselves unfamiliar. Repeating foreign words containing unfamiliar phonemes activated the putamen to an even greater degree than did repeating native-language pseudo-words, and this difference decreased with training. The results of Simmonds et al. (2014) suggest that the role of the putamen may not be restricted to producing novel motor sequences, but may also relate to producing novel articulatory movements. The current study extended these findings from speech to song and provided specific evidence that the putamen has a role in vocal imitation beyond the auditory-perception or motor-execution components shared with

the two control conditions.

Figure 6 attempts to summarize the results of the present study in the context of the standard model of vocal imitation in human neuroscience literature, namely the Wernicke-Geschwind model, which emphasizes the transmission of auditory information from the superior temporal gyrus to the inferior frontal gyrus via the arcuate fasciculus. We argue that this pathway is necessary but not sufficient for vocal imitation to occur. Instead, critical processes in the putamen beyond those required for either stimulus-encoding or production alone are needed in order to match target sounds to vocal motor commands. At present, it is uncertain if the critical connectivity between the basal ganglia and the vocal-motor system occurs with the IFG, larynx motor cortex (via the SMA), or both. Future studies of both functional and structural connectivity will be needed to resolve this issue.

4.5.1 Evolutionary considerations

Comparative neuroscience has revealed evolutionary expansions of brain regions throughout the human audio-vocal system relative to other primates, which has generated several neuroanatomical hypotheses for the evolution of vocal imitation. However, the evolution of vocal imitation is phylogenetically coupled with flexible motor control over the vocal organ, be it a larynx or a syrinx. We are not aware of any species that has the capacity to flexibly produce novel vocalizations in the absence vocal imitation, or vice versa. Hence, while undoubtedly useful, anatomical comparisons between species necessarily confound adaptations that underlie the sensorimotor transformations required

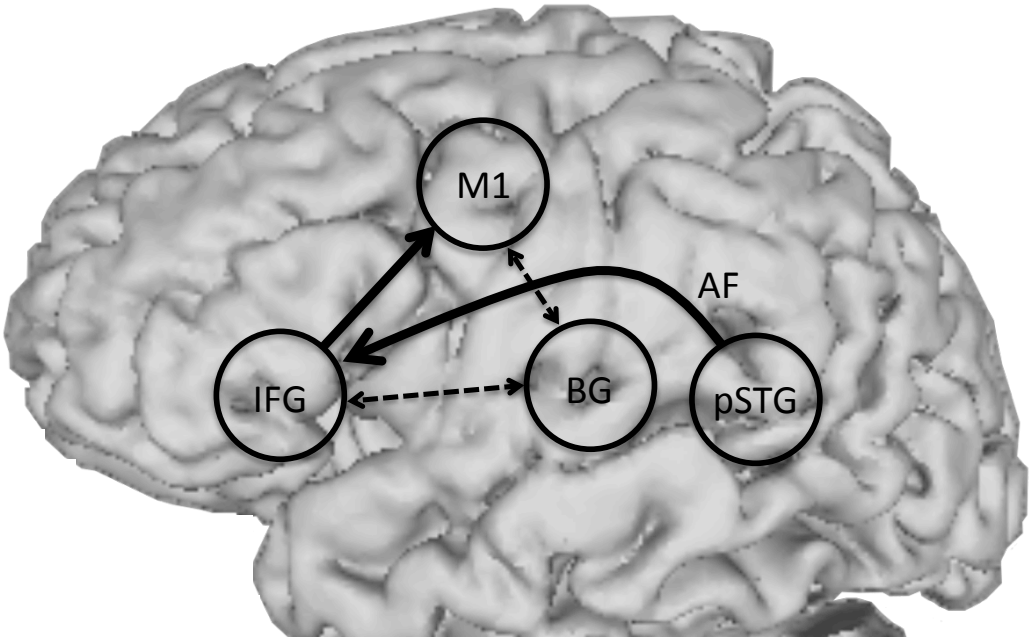

Figure 4.6: A simple neural model of vocal imitation.

The model summarizes the results of the present study in the context of the standard model of vocal imitation in human neuroscience literature, namely the Wernicke-Geschwind model. Target sounds are processed in auditory regions, including the posterior part of the superior temporal gyrus (STG), and are transmitted to the frontal lobe along the arcuate fasciculus (AF) to the inferior frontal gyrus (IFG), which in turn projects to the primary motor cortex, which executes motor commands to reproduce the target sound. Results from the current study suggest that processing through the cortico-striate loop is necessary for matching auditory targets with motor commands. However, it is unclear both from the present experiment and from songbird models of this system whether the critical anatomical connection with the basal ganglia occurs at the level of the inferior frontal gyrus or motor cortex. This uncertainty is indicated by the dashed lines connecting these structures to the basal ganglia.

for vocal imitation with sensory or motor adaptations that underlie the capacity for flexible control of the vocal organ. What does seem clear, however, is that the human audio-vocal system evolved the capacity to perform vocal imitation from phylogenetic precursors that lacked both of these abilities.

Several neuro-phenotypical differences have been described between humans and other primates that may be relevant to the emergence of vocal imitation, flexible vocal control, or both. In humans, the arcuate fasciculus is more strongly developed than in non-human apes (Rilling et al., 2008; Rilling, Glasser, Jbabdi, Andersson, & Preuss, 2012). The IFG pars opercularis contains the evolutionarily-novel diagonal sulcus, which is associated with increased cortical volume of this area (Keller, Roberts, & Hopkins, 2009). In humans, corticobulbar neurons from the motor cortex project directly to the nucleus ambiguus (Iwatsubo, Kuzuhara, & Kanemitsu, 1990; Kuypers, 1958b), while such direct connections are sparse in chimpanzees (Kuypers, 1958a) and absent in monkeys (Jürgens & Ehrenreich, 2007). In addition, the cortical larynx area has undergone an evolutionary migration from the premotor cortex in monkeys (Hast, Fischer, & Wetzel, 1974) to an intermediate position in great apes (Leyton & Sherrington, 1917) to the primary motor cortex in humans (Bouchard, Mesgarani, Johnson, & Chang, 2013; Brown, Ngan, & Liotti, 2008; Loucks et al., 2007; Pfenning et al., 2014). While comparative neuroscience has greatly advanced our knowledge of brain evolution, such neuroanatomical differences cannot be specifically attributed to the emergence of vocal imitation in humans without further functional evidence.

Some of the critical evidence that comes to bear on the evolution of the

vocal system comes not from a consideration of homology with primates but of analogy with other vocal-learning species, most notably songbirds. A large body of evidence links songbird Area X − which is a vocally-specialized region of the striatum − to imitation (Jarvis, 2007). Furthermore, there are marked anatomical and molecular similarities between the human and songbird vocal systems, which may reflect a process of convergent evolution (Jarvis, 2007; Petkov & Jarvis, 2012; Pfenning et al., 2014). Lesions to Area X and related structures disrupt vocal learning, but have little effect on pre-learned song (Bottjer et al., 1984; Sohrabji et al., 1990). These structures contain neurons that may compute causal inverse models that relate target sounds to motor commands (Giret, Kornfeld, Ganguli, & Hahnloser, 2014). Causal inverse models are maximally efficient for motor learning if they generate variable motor commands (Hanuschkin, Ganguli, & Hahnloser, 2013), since variability is required for motor exploration and thus for improvement on subsequent imitative attempts. Ablating output from the forebrain-striatal loop, such that only the posterior descending pathway drives vocalization, results in highly stereotyped song. In contrast, ablating part of the descending pathway, such that only the forebrain-striatal loop drives vocalization, results in a reversion to the oscine equivalent of babbling, which is characterized by highly variable song (Aronov, Andalman, & Fee, 2008). Song is typically more variable during undirected singing than when it is directed from a male to a female. Increased variability in neural firing along the forebrain-striatal loop during undirected singing (Hessler & Doupe, 1999) results in increased song variability (Kao, Doupe, & Brainard, 2005; Liu & Nottebohm, 2005), and lesioning this pathway

prevents such context-dependent changes in song variability to occur (Kao & Brainard, 2006). The forebrain striatal-loop is therefore believed to participate in both generating causal inverse models to produce new motor programs and in modulating motor variability to facilitate motor exploration and learning.

One gene that links the vocal systems of humans and songbirds is FOXP2. Experimental knockdown of FOXP2 in the juvenile songbird's Area X selectively disrupts vocal imitation (Haesler et al., 2007), and FOXP2 expression in this region continues to modulate song variability into adulthood (Teramitsu & White, 2006). In humans, FOXP2 mutations are associated with extensive speech and language deficits (Hurst, Baraitser, Auger, Graham, & Norell, 1990; Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001), including the inability to imitate novel speech sounds, such as pseudo-words (Shriberg et al., 2006; Watkins, Dronkers, & Vargha-Khadem, 2002). Patients with FOXP2 mutations have reduced activation throughout the vocal system, including the putamen, during pseudo-word repetition tasks (Liégeois, Morgan, Connelly, & Vargha-Khadem, 2011). The existing literature is broadly consistent with an analogous role of FOXP2 in humans and songbirds. However, such a conclusion is limited by the necessary reliance on natural experiments in humans. Experimental evidence from the current study further supports the functional analogy between the songbird forebrain-striatal loop and the human cortico-striate loop by demonstrating for the first time that the human putamen is preferentially activated during vocal pitch imitation compared to a well-matched non-imitative vocalization task.

The current study demonstrated that, in humans, the putamen is preferentially engaged by vocal imitation, but it is by no means exclusive to imitative processes. This might suggest a potential species difference between humans and songbirds. Indeed, lesions of Area X in songbirds are not believed to impair the production of songs that have already been learned (although see Hessler & Doupe, 1999; Kao & Brainard, 2006; Kubikova et al., 2014), whereas disruption of the basal ganglia system in humans leads to strong vocal-production deficits. Degenerative diseases of the basal ganglia, such as Parkinson's disease, can cause severe forms of dysphonia and articulatory disturbances (Blumin, Pcolinsky, & Atkins, 2004; Canter, 1963). This suggests that, as with basal ganglia control of other effectors, the vocal portion of the putamen supports vocal production. The putamen also co-activates with the rest of the vocal system both when vocalizing (Brown et al., 2009) and when discriminating pitch patterns (Brown & Martinez, 2007). This suggests that the basal ganglia may have an underappreciated role in non-motor functions (Kotz, Schwartze, & Schmidt-Kassow, 2009).

The position of the putamen within the human vocal system remains unclear. In songbirds, Area X receives input from a region whose hypothesized human analogue is Broca's area (Petkov & Jarvis, 2012). However, evidence for this analogy remains sparse (Pfenning et al., 2014). Alternatively, the human vocal striatum may receive projections from the SMA, which is the dominant source of afferent fibers for cortico-striate motor loops supporting other effectors (Alexander, DeLong, & Strick, 1986; Kunzle, 1975). Indeed, in the present study, the SMA, and not the IFG, was preferentially engaged

by vocal imitation, suggesting that the SMA may be linked with the putamen during vocal imitation. However, diffusion tensor imaging of the human brain suggests that both the IFG (Ford et al., 2013) and SMA project to the putamen (Leh, Ptito, Chakravarty, & Strafella, 2007; Lehéricy et al., 2004). Further research is required to elucidate the anatomical and functional connectivity of the putamen within the vocal motor system.

4.6 Conclusions

We report the results of a highly controlled brain imaging study of vocal pitch imitation in humans. While the tasks of imitating a novel melody and singing a familiar melody from memory both robustly activated a network of vocal areas, imitation was associated with greater activation in a subset of this network, most prominently the putamen. This region is the putative analogue of a critical node in the forebrain-striatal loop for vocal learning in songbirds. These data provide the first evidence that the putamen − but not Broca's Area − is preferentially engaged during imitative singing in humans, as predicted by functional analogy with songbird Area X.

4.7 Acknowledgements

This work was funded by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to S.B (grant no. 371336).

4.8 References

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience, 9*, 357−381.

Aronov, D., Andalman, A. S., & Fee, M. S. (2008). A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science, 320*(5876), 630−634.

Arriaga, G., & Jarvis, E. D. (2013). Mouse vocal communication system: Are ultrasounds learned or innate? *Brain and Language, 124*(1), 96−116.

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology, 5*(1), 69−83.

Blumin, J. H., Pcolinsky, D. E., & Atkins, J. P. (2004). Laryngeal findings in advanced Parkinson's disease. *Annals of Otology, Rhinology and Laryngology, 113*, 253−258.

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer [Computer program]. Version 5.1.29, retrieved 11 March 2010 from http://www.praat.org/

Bottjer, S., Miesner, E., & Arnold, A. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science, 224*(4651), 901−903.

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature, 495*(21), 327−332.

Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Turkeltaub, P., & Liotti, M. (2009). The somatotopy of speech: Phonation and articulation in the human motor cortex. *Brain and Cognition, 70*(1), 31−41.

Brown, S., & Martinez, M. J. (2007). Activation of premotor vocal areas during musical discrimination. *Brain and Cognition, 63*, 59−69.

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. *Cognitive Brain Research,*

20(3), 363−375.

Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex, 18*(4), 837−845.

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., & Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - An aggregate analysis of lesion and fMRI data. *Brain and Language, 119*(3), 119−128.

Canter, G. J. (1963). Speech characteristics of patients with Parkinson's disease: I. intensity, pitch and duration. *Journal of Speech and Hearing Disorders, 28*, 221−229.

Chang, S. E., Kenney, M. K., Loucks, T. M. J., Poletto, C. J., & Ludlow, C. L. (2009). Common neural substrates support speech and non-speech vocal tract gestures. *Neuroimage, 47*(1), 314−325.

Dalla Bella, S., & Berkowska, M. (2009). Singing proficiency in the majority: Normality and "phenotypes" of poor singing. *Annals of the New York Academy of Sciences, 1169*, 99−107.

Damasio, H., & Damasio, A. R. (1980). The anatomical basis of prosopagnosia. *Brain, 103*, 337−350.

Di Pallegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study.*Experimental Brain Research, 91*, 176−180.

Ford, A. A., Triplett, W., Sudhyadhom, A., Gullett, J., McGregor, K., Fitzgerald, D. B., et al. (2013). Broca's area and its striatal and thalamic connections: A diffusion-MRI tractography study. *Frontiers in Neuroanatomy, 7*, 1−8.

Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: Single-subject analyses of unsmoothed fMRI data. *Cerebral Cortex, 19*(6), 1239−1255.

Geschwind, N. (1970). The organization of language and the brain. *Science, 170*(3961), 940−944.

Giret, N., Kornfeld, J., Ganguli, S., & Hahnloser, R. H. R. (2014). Evidence for a causal inverse model in an avian cortico-basal ganglia circuit. *Proceedings of the National Academy of Sciences of the United States of America, 111*(16), 6063−6068.

Grillner, S., Robertson, B., & Stephenson-Jones, M. (2013). The evolutionary origin of the vertebrate basal ganglia and its role in action selection. *The Journal of Physiology*, 591(2011), 5425−5431.

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biology, 5*(12), e321.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A., Summerfield, A. Q., Elliott, M. R., et al. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping, 7*(3), 213−223.

Hanuschkin, A., Ganguli, S., & Hahnloser, R. H. R. (2013). A Hebbian learning rule gives rise to mirror neurons and links them to control theoretic inverse models. *Frontiers in Neural Circuits*, 7, 106.

Hast, M. H., Fischer, J. M., & Wetzel, A. B. (1974). Cortical motor representation of the laryngeal muscles in macaca mulatta. *Brain, 73*, 229−240.

Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., & Mazziotta, J. C. (2003). The essential role of Broca's area in imitation. *European Journal of Neuroscience, 17*(5), 1123−1128.

Hessler, N. A., & Doupe, A. J. (1999). Social context modulates singing-related neural activity in the songbird forebrain. *Nature Neuroscience, 2*(3), 209−211.

Hope, T. M. H., Prejawa, S., Jones, O. P., Oberhuber, M., Seghier, M. L., Green, D. W., & Price, C. J. (2014). Dissecting the functional anatomy of auditory word repetition. *Frontiers in Human Neuroscience, 8*, 1−17.

Hurst, J. A., Baraitser, M., Auger, E., Graham, F., & Norell, S. (1990). An extended family with a dominantly inherited speech disorder. *Developmental Medicine and Child Neurology, 32*, 352−355.

Hutchins, S., & Moreno, S. (2013). The Linked Dual Representation model of vocal perception and production. *Frontiers in Psychology, 4*, 825.

Iacoboni, M., Woods, R., Brass, M., Bekkering, H., Mazziotta, J., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*, 2526−2528.

Iwatsubo, T., Kuzuhara, S., & Kanemitsu, A. (1990). Corticofugal projections to the motor nuclei of the brainstem and spinal cord in humans. *Neurology, 40*, 309−312.

Jarvis, E. (2007). Neural systems for vocal learning in birds and humans: A synopsis. *Journal of Ornithology, 148*(1), 35−44.

Jarvis, E., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., et al. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience, 6*, 151−159.

Jürgens, U., & Ehrenreich, L. (2007). The descending motorcortical pathway to the laryngeal motoneurons in the squirrel monkey. *Brain Research, 1148*, 90−95.

Kao, M. H., & Brainard, M. S. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *Journal of Neurophysiology, 96*(3), 1441−1455.

Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005). Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature, 433*(7026), 638−643.

Keller, S. S., Roberts, N., & Hopkins, W. (2009). A comparative magnetic resonance imaging study of the anatomy, variability, and asymmetry of Brocas area in the human and chimpanzee brain. *Journal of Neuroscience, 29*(46), 14607−14616.

King, S., & Sayigh, L. (2013). Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proceedings of the Royal Society B, 280*, 1−9.

Knörnschild, M., Nagy, M., Metz, M., Mayer, F., & von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biology Letters, 6*(2),

156−159.

Kotz, S. A., Schwartze, M., & Schmidt-Kassow, M. (2009). Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. *Cortex, 45*(8), 982−990.

Kubikova, L., Bosikova, E., Cvikova, M., Lukacova, K., Scharff, C., & Jarvis, E. D. (2014). Basal ganglia function, stuttering, sequencing, and repair in adult songbirds. *Scientific Reports, 4*, 6590.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation. *Journal of the Acoustical Society of America, 100*(4), 2425−2438.

Kunzle, H. (1975). Bilateral projections from precentral motor cortex to the putamen and other parts of the basal ganglia. An autoradiographic study in Macaca fascicularis. *Brain Research, 88*(1975), 195−209.

Kuypers, H. G. J. M. (1958a). Corticobulbar connexions to the pons and lower brain-stem in man. *Brain, 81*(3), 364−388.

Kuypers, H. G. J. M. (1958b). Some projections from the peri-central cortex to the pons and lower brain stem in monkey and chimpanzee. *Journal of Comparative Neurology, 110*, 221−255.

Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413*(6855), 519−523.

Leh, S. E., Ptito, A., Chakravarty, M. M., & Strafella, A. P. (2007). Fronto-striatal connections in the human brain: A probabilistic diffusion tractography study. *Neuroscience Letters, 419*(2), 113−118.

Lehéricy, S., Ducros, M., Krainik, A., Francois, C., Van De Moortele, P. F., Ugurbil, K., & Kim, D. S. (2004). 3-D diffusion tensor axonal tracking shows distinct SMA and pre-SMA projections to the human striatum. *Cerebral Cortex, 14*, 1302−1309.

Leyton, S., & Sherrington, C. (1917). Observations on the excitable cortex of the chimpanzee, organgutan, and gorilla. *Experimental Physiology, 11*(2),

135−222.

Liégeois, F., Morgan, A. T., Connelly, A., & Vargha-Khadem, F. (2011). Endophenotypes of FOXP2: Dysfunction within the human articulatory network. *European Journal of Paediatric Neurology, 15*(4), 283−288.

Liu, W., & Nottebohm, F. (2005). Variable rate of singing and variable song duration are associated with high immediate early gene expression in two anterior forebrain song nuclei. *Proceedings of the National Academy of Sciences, 102*(30), 10724−10729.

Loucks, T. M. J., Poletto, C. J., Simonyan, K., Reynolds, C. L., & Ludlow, C. L. (2007). Human brain activation during phonation and exhalation: Common volitional control for two upper airway functions. *Neuroimage, 36*(1), 131−43.

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2009). Is the mirror neuron system involved in imitation? A short review and meta-analysis. *Neuroscience and Biobehavioral Reviews, 33*(7), 975−980.

Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., & Jenner, K. C. (2000). Cultural revolution in whale songs. *Nature, 408*(6812), 537.

Nottebohm, F. (1972). The origins of vocal learning. *American Naturalist, 106*(947), 116−140.

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, Serinus canarius. *Journal of Comparative Neurology*, 165(4), 457−486.

Papousek, M. (1996). Intuitive parenting: A hidden source of musical stimulation in infancy. In I. Deliège& J. Sloboda (Eds.), *Musical beginnings: Origins and development of musical competence* (pp. 88−112). Oxford: Oxford University Press.

Papoutsi, M., de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: An fMRI study of the role of Brocas area in speech production. *Cerebral Cortex, 19*(9), 2156−2165.

Peeva, M. G., Guenther, F. H., Tourville, J. A., Nieto-Castanon, A., Anton, J.

L., Nazarian, B., & Alario, F. X. (2010). Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage, 50*(2), 626−638.

Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience, 4*, 1−24.

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science, 346*(6215), 1256846−1256846.

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of tone deafness. *Music Perception, 25*(2), 95−115.

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *Journal of the Acoustical Society of America, 128*(4), 2182−2190.

Pfordresher, P. Q., Demorest, S. M., Dalla Bella, S., Hutchins, S., Loui, P., Rutkowski, J., & Welch, G. F. (2015). Theoretical perspectives on singing accuracy: An introduction to the special issue on singing accuracy (part 1). *Music Perception, 32*, 227−231.

Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive Psychology, 70*, 31−57.

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Elephants are capable of vocal learning. *Nature, 434*, 455−456.

Poulson, C. L., Kymissis, E., Reeve, K. F., Andreators, M., & Reeve, L. (1991). Generalized vocal imitation in infants. Journal of *Experimental Child Psychology, 5*1, 267−279.

Prather, J. F., Peters, S., Nowicki, S., & Mooney, R. (2008). Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature, 451*(7176), 305−310.

Ralls, K., Fiorelli, P., & Gish, S. (1985). Vocalizations and vocal mimicry

in captive harbor seals, Phoca vitulina. *Canadian Journal of Zoology*, 63, 1050−1056.

Rauschecker, A. M., Pringle, A., & Watkins, K. E. (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Human Brain Mapping, 29*(11), 1231−1242.

Rilling, J. K., Glasser, M. F., Jbabdi, S., Andersson, J., & Preuss, T. M. (2012). Continuity, divergence, and the evolution of brain language pathways. *Frontiers in Evolutionary Neuroscience, 3*, 1−6.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., & Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience, 11*(4), 426−428.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131−141.

R Core Development Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Sanvito, S., Galimberti, F., & Miller, E. H. (2007). Observational evidences of vocal learning in southern elephant seals: A longitudinal study. *Ethology, 113*(2), 137−146.

Shmuelof, L., & Krakauer, J. W. (2011). Are we ready for a natural history of motor learning? *Neuron, 72*(3), 469−476.

Shriberg, L. D., Ballard, K. J., Tomblin, J. B., Duffy, J. R., Odell, K. H., & Williams, C. A. (2006). Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. *Journal of Speech, Language, and Hearing Research, 49*, 500−525.

Simmonds, A. J., Leech, R., Iverson, P., & Wise, R. J. S. (2014). The response of the anterior striatum during adult human vocal learning. Journal of Neurophysiology, 112, 792−801.

Sohrabji, F., Nordeen, E. J., & Nordeen, K. W. (1990). Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. *Behavioral and Neural Biology, 53*(1), 51−63.

Stoeger, A. S., Mietchen, D., Oh, S., de Silva, S., Herbst, C. T., Kwon, S., & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology, 22*(22), 2144−2148.

Studdert-Kennedy, M. (2000). Imitation and the emergence of segments. *Phonetica, 57*, 275−283.

Teramitsu, I., & White, S. A. (2006). FoxP2 regulation during undirected singing in adult songbirds. *Journal of Neuroscience, 26*(28), 7390−7394.

Trehub, S. E. (2001). Musical predispositions in infancy. In R. J. Zatorre& I. Peretz (Eds.), *The biological foundations of music* (pp. 1−16). New York: New York Academy of Sciences.

Watkins, K. E., Dronkers, N. F., & Vargha-Khadem, F. (2002). Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain, 125*o, 452−464.

Chapter 5

Stuttering as a trait or state – an ALE meta-analysis of neuroimaging studies

Michel Belyk, Shelly Jo Kraft & Steven Brown.

European Journal of Neuroscience, 2015, 41, 275-284.

5.1 Abstract

Stuttering is a speech disorder characterised by repetitions, prolongations and blocks that disrupt the forward movement of speech. An earlier meta-analysis of brain imaging studies of stuttering (Brown et al., 2005) revealed a general trend towards rightward lateralization of brain activations and hyperactivity in the larynx motor cortex bilaterally. The present study sought not only to update that meta-analysis with recent work but to introduce an important

distinction not present in the first study, namely the difference between "trait" and "state" stuttering. The analysis of trait stuttering compares people who stutter (PWS) with people who do not stutter when behaviour is controlled for, i.e., when speech is fluent in both groups. In contrast, the analysis of state stuttering examines PWS during episodes of stuttered speech compared with episodes of fluent speech. Seventeen studies were analysed using activation likelihood estimation. Trait stuttering was characterised by the well-known rightward shift in lateralization for language and speech areas. State stuttering revealed a more diverse pattern. Abnormal activation of larynx and lip motor cortex was common to the two analyses. State stuttering was associated with overactivation in the right hemisphere larynx and lip motor cortex. Trait stuttering was associated with overactivation of lip motor cortex in the right hemisphere but underactivation of larynx motor cortex in the left hemisphere. These results support a large literature highlighting laryngeal and lip involvement in the symptomatology of stuttering, and disambiguate two possible sources of activation in neuroimaging studies of persistent developmental

5.2 Introduction

Stuttering is a disorder characterised by speech with involuntary repetitions, prolongations, hesitations and blocks at the levels of syllables and words (Wingate, 1964). Theories of stuttering attribute its etiology to a wide variety of factors, including disordered sensory feedback (Max et al., 2004), linguistic deficits (Postma & Kolk, 1993; Howell, 2004), anticipation of speech difficulties (Brocklehurst et al., 2013), generalised motor deficits (Forster &

Webster, 2001) and/or speech-specific motor deficits (Namasivayam & van Lieshout, 2011), including a strong genetic influence (see review by Kraft & Yairi, 2012). An activation likelihood estimation (ALE) meta-analysis of the neuroimaging literature on persistent developmental stuttering (Brown et al., 2005) provided support for a diversity of underlying mechanisms, including overactivation of motor areas, underactivation of auditory areas, and anomalous right-hemisphere activation in regions not seen in fluent individuals.

However, research on stuttering frequently distinguishes between the person who stutters (i.e., "trait" stuttering) and the act of stuttering (i.e., "state" stuttering). An important question that comes from the observation of activation differences between people who stutter (PWS) and people who do not stutter (PWNS) is whether these differences are episodic, i.e., occurring only during bouts of stuttering, or whether they are stable features of the brains of PWS.

Stuttering is characterised not only by a propensity to produce stuttered speech but by abnormalities in speech motor control (Namasivayam & van Lieshout, 2011), non-speech motor skills (Neef et al., 2011a), auditory-processing abilities (Toscher & Rupp, 1978) and possibly language abilities (Ntourou et al., 2011; although see Nippold, 2012 for a refutation of this association). However, PWS do not always stutter. Stuttering occurs episodically, and the fluency state of a person who stutters is modulated by a broad array of contextual factors. PWS stutter more when speech difficulty is anticipated (Rappaport & Bloodstein, 1971), when using contrastive stress (Klouda &

Cooper, 1988), and at the onset of voicing (Adams & Reis, 1971). PWS stutter less when they speak in a whisper (Commodore & Cooper, 1978), speak quietly, slowly (Johnson & Rosen, 1937) or with prior rehearsal (Brenner et al., 1972), when they sing, speak rhythmically to a metronome or in chorus with a recording of the text they are reading aloud (Davidow et al., 2009), when they speak in the presence of auditory noise (Garber & Martin, 1977), or when auditory feedback is altered (Stuart et al., 1997). Stuttering, therefore, presents the paradoxical picture that, while a propensity to stutter is a relatively constant trait, a person's state of fluency can be modulated by a host of contextual factors that can provide immediate, although transient, remediation from stuttering. We carried out an updated ALE meta-analysis of the neuroimaging literature on developmental stuttering that incorporated this important trait−state distinction. In particular, the analysis of trait stuttering compares PWS with PWNS when behaviour is controlled for, i.e., when speech is fluent in both groups. In contrast, the analysis of state stuttering examines PWS during episodes of stuttered speech compared with episodes of fluent speech.

5.3 Methods

Activation likelihood estimation is a meta-analytic technique for ascertaining the regions of concordant activation across a corpus of brain imaging studies (Turkeltaub et al., 2002). Each activation focus is modeled as a three-dimensional Gaussian probability distribution whose width is determined by the size of the subject group so as to reflect increasing certainty with increasing

sample size (Eickhoff et al., 2009). Maps of activation likelihoods are created for each study by taking the maximum probability of activation at each voxel. A random-effects analysis then tests for the convergence of activations across studies vis-a-vis a null hypothesis of spatially independent brain activations.

5.3.1 General inclusion criteria

Published studies were searched using the Web of Knowledge and Pubmed databases with the search terms "stuttering + fMRI" and "stuttering + PET", where fMRI refers to functional magnetic resonance imaging and PET refers to positron emission tomography. The reference sections of the retrieved publications were searched for additional studies. To be included in the meta-analyses, studies had to (i) be published in a peer-reviewed scientific journal, (ii) report coordinate-based analyses of the data in a standard stereotaxic space, (iii) image the whole brain or nearly the whole brain, (iv) scan developmental stutterers and (v) have subjects perform overt speech tasks. The search returned 34 publications, 24 of which met our inclusion criteria. Several of the remaining articles reported previously published data and therefore did not contribute independent results to the data set. These data were combined according to subject group, as recommended by Turkeltaub et al. (2011). One study reported data for individual subjects but no group-level analysis (Wymbs et al., 2013). Single-subject data were treated as individual experiments with n = 1. The present analysis included data from 21 unique subject groups reported across 18 publications, totaling 213 PWS and 186 PWNS. In all but one study, participants were audio-recorded during speech tasks in the

scanner. The exception was Howell et al. (2012) who nonetheless determined the absence of stuttering by ear. MNI coordinates were transformed to Talairach coordinates (Talairach & Tournoux, 1988). ALE analyses were carried out using GingerALE 2.3, employing the False Discovery Rate correction for multiple comparisons ($P < 0.01$), with a cluster threshold $k > 10$.

5.3.2 Trait vs. state stuttering

Experiments were subdivided into those that examined stuttering as a stable trait and those that examined it as an episodic state (Fig. 1). The meta-analysis of trait stuttering included contrasts between PWS and PWNS when both spoke fluently. To be included in the analysis of trait stuttering, studies had to pass two additional criteria: (i) they had to confirm that all participants spoke fluently during image collection and (ii) they had to report either direct contrasts between brain images of PWS while they spoke fluently vs. PWNS speaking under matched conditions, or report correlations between brain imaging data and stuttering severity, as measured outside the scanner ($n = 8$ for overactivation and $n = 9$ for underactivation relative to PWNS). These studies reveal stable neural features of PWS during fluent speech.

The meta-analysis of state stuttering included contrasts exclusively for PWS, and examined when PWS stuttered compared to when they spoke fluently. To be included in the analysis of state stuttering, studies had to pass two additional criteria: (i) confirm that PWS stuttered during image collection in stuttering conditions but not in fluent conditions and (ii) report direct contrasts between brain images while PWS stuttered vs. when they spoke

fluently, or report correlations between brain images and rates of stuttering in the scanner. All state-stuttering contrasts included one scan in which PWS stuttered, although some of the earlier studies contained stutters embedded in otherwise fluent speech. Those studies might best be described as capturing speech that is prone to stuttering rather than stuttering per se. Nonetheless, to the extent that those stuttering scans were diluted by fluent speech, comparisons between these scans and scans containing only fluent speech are conservative in that they should underestimate differences between stuttered and fluent speech production. Table 1 lists the tasks performed in each study and the rate of stuttering for each study where applicable. These studies reveal the neural features associated with episodes of stuttering (n = 10 for overactivation and n = 8 for underactivation relative to fluent speech).

The analyses of both trait and state stuttering included brain-imaging data on PWS while they spoke fluently (see scheme in Fig. 1). In two studies contributing to these analyses, fluency was achieved by instructing participants to speak with a metronome (Braun et al., 1997; Toyomura et al., 2011), speak over-learned content (Braun et al., 1997), or speak in chorus with another speaker (Toyomura et al., 2011), all of which facilitated fluency (see Table 1). In these cases, experimental conditions and behaviour were matched between PWS and PWNS. However, in most of the studies contributing to our analyses, participants were spontaneously fluent. Spontaneous fluency may occur when speech tasks are restricted to short utterances or when scanner noise facilitates fluency. Some studies made no attempt to manipulate stuttering, but instead classified utterances as stuttered or fluent following data collection (i.e., den

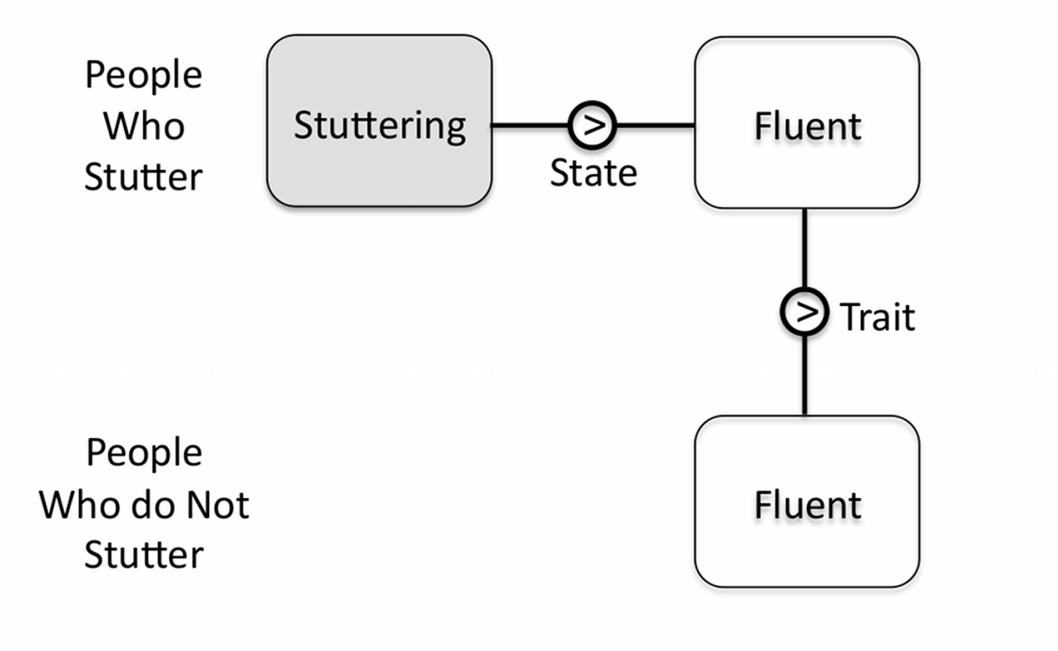

Figure 5.1: Operational definitions of trait and state stuttering.

Trait stuttering is revealed by contrasts between PWS and PWNS during fluent speech (PWS fluent > PWNS fluent). It is a between-group comparison. State stuttering is revealed by contrasts within PWS during stuttered vs. fluent speech (PWS stuttering > PWS fluent). It is a within-group comparison.

Ouden et al., 2013; Wymbs et al., 2013).

The analyses of trait and state stuttering have a parallel structure in that both are based on contrasts with PWS while speaking fluently, as shown graphically in Fig. 1. In addition, the analysis of state stuttering may be interpreted as being additive with the analysis of trait stuttering. Trait stuttering reflects the fluent speech of PWS relative to PWNS and thus represents the background condition of PWS. State stuttering, then, reflects additional changes beyond that background state that occur during episodes of stuttering.

5.4 Results

Figure 2 presents the ALE results on axial slices, and Table 2 provides Talairach coordinates for the ALE foci. We examine trait and state stuttering in sequence. Trait stuttering showed increased likelihood of activation mainly in the right hemisphere, supportive of classic right-shift models of stuttering. The right hemisphere homologue of Broca's area, specifically the inferior frontal gyrus (IFG) pars opercularis or Brodmann Area (BA) 44, was more active in the brains of PWS than PWNS during fluent speech, as were other right-hemisphere premotor areas, including the presupplementary motor area (SMA), lateral premotor cortex in the precentral gyrus (BA 6), lip motor cortex (BA 4/6) and Rolandic operculum. Similar trends were observed in the right IFG pars orbitalis extending into the ventral insula, superior frontal gyrus (BA 6), superior frontal gyrus (BA 9), inferior parietal lobule (BA 40) and bilateral superior parietal lobule (BA 7).

Table 5.1: Summary of studies included in the meta-analyses.

This table lists the seventeen studies that contributed to the analyses of trait and/or state stuttering. Studies contributed to a positive association with trait stuttering if they reported the directional contrast [PWSfluent > PWNSfluent] and to a positive association with state stuttering if they reported the directional contrast [PWSstuttering > PWNSfluent]. Studies that reported contrasts in the opposite direction contributed to negative associations. The stuttering and fluent tasks columns list the tasks performed in each study. Rates of stuttering are reported in parentheses where applicable. The final column identifies studies that reported correlation with either (trait) severity of stuttering or (state) stuttering rate in the scanner in addition to or instead of high level contrasts. *Stuttering confirmed, but rates not reported. †Percentage of 4-s intervals which contained stuttering. ⋄Percentage of Japanese morea which were stuttered. ‡Percentage of utterances which were stuttered.

	trait		state				
Study	+	-	+	-	Dysfluent Task	Fluent Task	Correlation
Trait stuttering							
Braun et al. (1997)	X	X			-	Overlearned, paced speech	
Neumann et al. (2003)	X	X			-	Reading short sentences	
Priebishch et al. (2003)	X	X			-	Reading short sentences	
Giraud et al. (2008)	X	X			-	Reading short sentences	Severity
De Nil et al. (2003)	X	X			-	Word repetition	
Chang et al. (2009)	X	X			-	Monosyllalbe repetition	
Kell et al. (2009)	X	X			-	Reading short sentences	
Sakai et al. (2009)		X			-	Reading short sentences	
Lu et al. (2010)	X	X			-	Reading single words	
Howell et al. (2012)	X	X			-	Reading single words	
State stuttering							
Braun et al. (1997)			X	X	Recount narrative/sentence generation*	Overlearned/paced speech	Rate
Fox et al. (2000)			X		Reading paragrpahs (62%)†	-	Rate
Ingham et al. (2004)			X	X	Reading paragrpahs (73%)†	-	Rate
Toyomura et al (2011)				X	Reading short sentences (2.5%)⋄	Chorus/paced speech	
Ingham et al. (2012)			X	X	Paragraphs/narrative (9.75/8.84%)‡	-	Rate
Jiang et al. (2012)			X	X	Sentence completion (100%)‡	Sentence completion	
Wymbs et al (2013)			X	X	Reading words (100%)‡	Reading words	
den Ouden et al. (2013)			X	X	Reading words (100%)‡	Reading words	

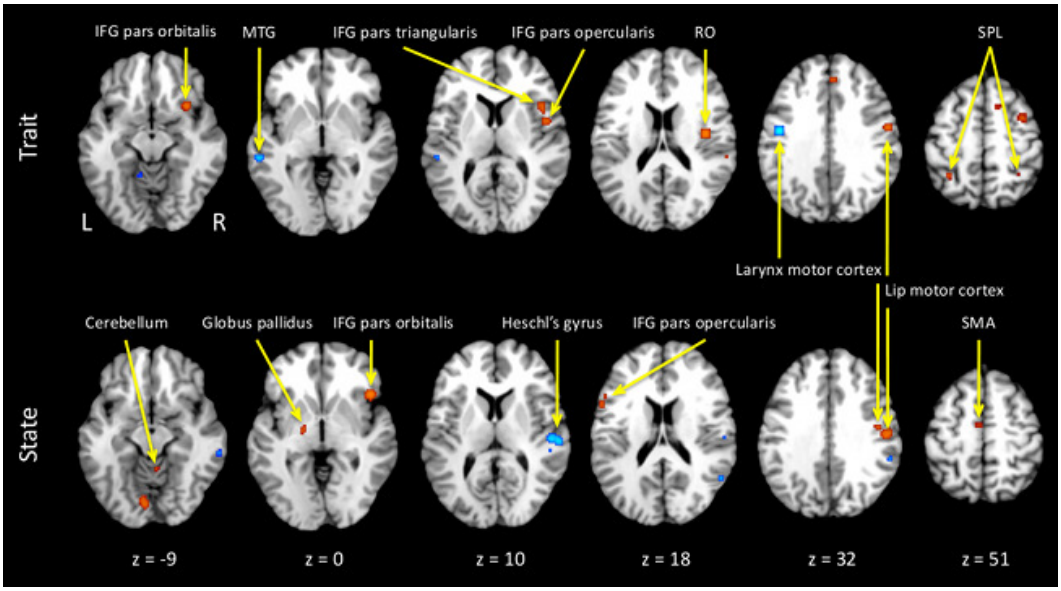

Figure 5.2: Results of the ALE analyses.

Axial slices in neurological convention showing regions consistently reported for trait and state stuttering. IFG, inferior frontal gyrus; MTG, middle temporal gyrus; RO, Rolandic operculum; SPL, superior parietal lobule; SMA, supplementary motor area.

Complementary to this result, trait stuttering showed decreased likelihood of activation exclusively in the left hemisphere. The most prominent decrease was seen in the left larynx motor cortex. Decreases were also observed in temporal lobe auditory areas, including the middle temporal gyrus (BA 21) and Heschl's gyrus (BA 41). Finally, a decrease was seen in the left cerebellar vermis. Overall, trait stuttering showed a strong right-shift pattern, with right-hemisphere increases and left-hemisphere decreases. Looking now to the brain activations associated with bouts of stuttering in PWS (Fig. 2, lower panel), the pattern was more diverse, showing effects in both hemispheres. State stuttering was associated with increased likelihood of activation in right larynx motor cortex and lip motor cortex (BA 4) in the homologous location to the left-hemisphere underactivation seen for trait stuttering. Increases were also seen in the left SMA/Pre-SMA (BA 6), globus pallidus, precuneus (BA 7), Broca's area corresponding to both the IFG pars opercularis (BA 44) and pars triangularis (BA 45), bilateral cerebellum and right IFG pars orbitalis (BA 47).

State stuttering showed decreased likelihood of activation exclusively in the right hemisphere. Most notably, decreases were observed in right hemisphere auditory areas, including Heschl's gyrus (BA 41), the posterior superior temporal gyrus (BA 22) and middle temporal gyrus (BA 21). Decreases were also observed in the supramarginal gyrus (BA 40) and middle frontal gyrus (BA 46).

In order to assess the reliability of the data, we determined hsy of the source studies reported activations in regions corresponding to each of the

ALE foci. The results are shown in Table 2 as the proportion of the source-studies contributing to each peak. Among the most reliable ALE foci were: trait overactivation of the right precentral gyrus (0.50) and Broca's homologue (0.38); trait underactivation of left larynx motor cortex (0.44); state overactivation of the SMA (0.60), lip motor cortex (0.50), cerebellar vermis (0.50) and IFG pars orbitalis (0.40); and state underactivation of right auditory cortex (0.50).

5.5 Discussion

In the present study, we expanded upon an earlier meta-analysis of brain imaging studies of stuttering (Brown et al., 2005) in order to examine the reliability of findings across the literature as well as to introduce a useful distinction not considered in the earlier analysis, namely that between stuttering as a stable trait and stuttering as a transient state. Overall, the findings were broadly consistent with the results of the earlier meta-analysis, showing overactivation in motor areas and underactivation in auditory areas. This argues for a general reliability of the findings of the last 10 years' worth of publications as well as a consistency in the analysis after switching to the more recent ALE methodology (Turkeltaub et al., 2011).

A novel approach of this study was to partition the meta-analysis results into trait vs. state effects. We assess trait stuttering by a between-group comparison during fluent speech. In contrast, we assess state stuttering by a within-group comparison, looking at fluent vs. stuttered speech. This basic distinction is pervasive in the literature that we have reviewed. However, the

literature must be interpreted with caution as it is unclear whether differences in brain activation are causes of stuttering or merely correlates. For example, studies of trait stuttering sometimes evoke fluency with task manipulations that may differentially affect activations in PWS and PWNS (although few studies that contributed to the current analysis did so). Furthermore, adult PWS have a lifetime of experience coping with their disorder. Trait stuttering could therefore reveal either brain abnormalities that cause the disorder or those that may compensate for it. Studies of children who stutter are few, but will be informative for this field (e.g., Chang et al., 2008). Similarly, studies of state stuttering may reveal causes of the stuttering event, attempts to compensate for stuttering or the correlates of stuttering as a motor act. Nonetheless, the trait−state distinction provides a useful disambiguation of the neural correlates of stuttering.

The previous meta-analysis by Brown et al. (2005) identified three "neural signatures' of stuttering, namely (i) overactivation of the right IFG/frontal operculum, (ii) underactivation of auditory cortex and (iii) overactivation of the cerebellar vermis. The current analysis elaborates on these findings by observing that (i) the right IFG/frontal operculum overactivation is restricted to trait stuttering, (ii) underactivation of auditory cortex is common to both trait and state stuttering and (iii) while the cerebellar vermis is overactivated during state stuttering, it is underactivated in trait stuttering, indicating that the relationship between the cerebellum and stuttering may be more complex than previously supposed. We further observed that the well-established right-shift for the brain activations of PWS (Travis, 1978; De Nil et al., 2000) was

more consistently found in the trait analysis than in the state analysis. Trait stuttering was associated with an increased likelihood of activation almost exclusively in the right hemisphere and a decreased likelihood of activation almost exclusively in the left hemisphere. The analysis of state stuttering, on the other hand, revealed increases in both hemispheres and decreases exclusively in the right hemisphere.

An important brain area that linked these two analyses was an area of primary motor cortex (x = 44, y = -8, z = 32) that matched the somatotopic location of the larynx motor cortex (x = 44, y = -10, z = 34) reported in Brown et al. (2008). State stuttering was associated with overactivation in the right hemisphere larynx motor cortex and trait stuttering with underactivation in the homologous region of the left hemisphere (see Fig. 2). The combination of the two results suggests a potential lack of coordination in the cortical control of the laryngeal muscles.

Before discussing the findings in detail, we would like to present a caveat. While an ALE meta-analysis detects brain regions that are commonly activated across studies, it is unable to detect differences between individuals. As previous research has demonstrated substantial individual differences in brain activations among PWS (Wymbs et al., 2013), we report a complementary analysis of the frequency of replication for each brain region that reached significance in our analysis (see Table 2). Hence we note that while our meta-analysis presents a unitary view of stuttering, PWS are a highly heterogeneous group of individuals, and stuttering as a syndrome may be comprised of multiple subtypes with distinct etiologies (Yairi, 2007). The results of individual

neuroimaging studies may therefore reflect mixtures of neural correlates from different subtypes of stuttering found within particular groups of subjects. The ALE method, in turn, identifies regions of the brain that are consistently reported as part of this mixture. Any individual person who stutters may manifest abnormal activation in only a subset of the regions we identified, in none of these regions, or in regions not reported here, in accordance with the etiology of their particular case.

Table 5.2: Cluster coordinates.The four sections of this table list brain regions that were either positively or negatively associated with trait and state stuttering. After each anatomical name in the brain region column, the Brodmann number for that region is listed. The columns labeled as x, y, and z contain the Talairach coordinates for the peak of each cluster. The mm^3 column lists the total volume of each cluster. "L" indicates local maxima contained within the region listed directly above. The ALE column lists the peak ALE estimate for each region multiplied by 10^3. The final column lists the proportion of studies contributing to each analysis that reported foci of activation directly in each brain region. IFG, inferior frontal gyrus; IPL, inferior parietal lobule; MTG, middle temporal gyrus; SMA, supplementary motor area; SMG, supramarginal gyrus; SPL, superior parietal lobule; STG, superior temporal gyrus.

Hemsiphere	Brain Region	Brodmann	x	y	z	mm^3	ALE (10^3)	Prop.
Trait stuttering								
Positive associations								
Right	Precentral gyrus	BA 6	36	6	50	192	10.73	0.50
Right	Lip motor cortex	BA 4/6	54	-4	30	296	14.21	0.38
Right	Rolandic operculum	BA 13	38	-10	18	496	16.12	0.38
Right	IFG pars opercularis	BA 44	48	2	8	232	13.23	0.38
Right	IFG pars opercularis	BA 44	42	16	12	288	11.71	0.38
Right	IPL	BA 40	58	-30	22	184	13.07	0.38
Right	SPL	BA 7	32	-46	46	168	12.38	0.25
Right	SMA	BA 6	14	14	54	192	11.58	0.25
Right	Medial frontal gyrus	BA 9	4	38	32	80	10.57	0.25
Right	IFG pars orbitalis	BA 47/13	32	16	-8	264	13.91	0.38
Left	SPL	BA 7	-30	-48	50	96	10.79	0.25
Negative associations								
Left	Larynx motor cortex	BA 4	-44	-8	32	376	14.19	0.44
Left	MTG	BA 21	-56	-32	-2	280	11.88	0.33
Left	Heschle's gyrus	BA 41	-52	-30	10	24	8.74	0.22
Left	Cerebellar vermis		-12	-48	-6	128	9.83	0.11
State stuttering								
Positive associations								

Continued on next page

Hemsiphere	Brain Region	Brodmann	x	y	z	mm^3	ALE (10^3)	Prop.
Right	SMA	BA 6	2	-22	58	16	8.44	0.60
Right	Lip motor cortex	BA 4	54	-14	34	696	13.06	0.50
Right	Larynx motor cortex	BA 4	44	-8	32	136	9.33	0.30
Right	IFG pars orbitalis	BA 47	44	20	0	520	13.86	0.40
Right	Cerebellar vermis		4	-46	-8	32	8.89	0.20
Left	SMA	BA 6	-4	-8	56	432	11.75	0.60
Left	SMA	BA 6	0	0	54	L	8.62	0.60
Left	Cerebellar vermis		-6	-78	-10	440	11.01	0.50
Left	IFG pars opercularis	BA 44	-56	12	16	160	9.45	0.30
Left	IFG pars triangularis	BA 45	-52	20	16	L	8.88	0.20
Left	Globus pallidus		-18	-12	0	80	9.11	0.20
Left	Precuneus	BA 7	-6	-56	48	16	8.43	0.20
Negative associations								
Right	Heschl's gyrus	BA 41	58	-22	8	1128	14.82	0.50
Right	Heschl's gyrus	BA 41	52	-20	8	L	12.57	0.50
Right	Heschl's gyrus	BA 41	56	-18	14	L	11.82	0.50
Right	Heschl's gyrus	BA 41	50	-28	14	200	11.57	0.50
Right	IPL	BA 40	38	-36	40	16	9.49	0.38
Right	Posterior STG	BA 22	52	-56	22	352	12.13	0.38
Right	SMG	BA 40	54	-38	32	24	9.23	0.38
Right	MTG	BA 21	62	-32	-10	72	9.9	0.25
Right	Middle frontal gyrus	BA 46	42	16	24	240	12.85	0.13

5.5.1 Basal ganglia and SMA

Alm (2004) proposed that dysfunction in the basal ganglia and corresponding cortical sites in the SMA may result in poor motor timing during speech production. Several neuroimaging studies have observed stuttering-related activation throughout the basal ganglia, including the caudate nucleus (Braun et al., 1997), putamen (Kell et al., 2009), globus pallidus (Ingham et al., 2004), subthalamic nucleus (Loucks et al., 2011), and substantia nigra (Wu et al., 1995). The meta-analysis of Brown et al. (2005) revealed the involvement of the SMA in stuttering but failed to detect any ALE foci in the basal ganglia. The present analysis suggests that state stuttering is associated with overactivation of the SMA while trait stuttering is associated with overactivation of the pre-SMA. This is consistent with the observation of increased activity of the orofacial muscles during stuttering, as the SMA has greater connectivity

with cortical motor areas (Luppino et al., 1993) and gives rise to more descending motor efferents (Dum & Strick, 1991) than does the preSMA. Both areas project to and receive projections from the basal ganglia (Inase et al., 1999; Akkal et al., 2007), although the present analysis observed overactivation in the globus pallidus of the basal ganglia for state stuttering only. However, this ALE focus was present in a relatively low proportion of studies, suggesting that this region is not reliably activated across studies of stuttering, as suggested previously by Brown et al. (2005). While it is clear that the basal ganglia are involved in stuttering, it is unclear as to which nucleus the abnormal activity is localised. The potential for a causal role of the basal ganglia in stuttering is highlighted by the case study of a recovered person who stutters and who underwent deep brain stimulation as treatment for Parkinson's disease (Burghaus et al., 2006). Stimulation of the subthalamic nucleus reduced Parkinsonian symptoms but caused a relapse of stuttering. Cessation of stimulation led to a return of Parkinsonian symptoms coupled with an abatement of stuttering. The pattern of brain activation in this subject in the presence vs. absence of deep brain stimulation revealed increased activation in many of the stutter-related regions reported in both Brown et al. (2005) and the current analysis, including the SMA, motor cortex and cerebellar vermis, as well as reduced activation in the auditory cortex.

5.5.2 Auditory cortex

There has been suggestive evidence since the earliest imaging studies of stuttering (Wu et al., 1995; Fox et al., 1996) that auditory areas are underactivated

in PWS. The Brown et al. (2005) meta-analysis failed to detect auditory deactivations in the subtraction between PWS and PWNS, although there was a trend for reduced activity in PWS in the auditory cortex bilaterally. The present meta-analysis was able to shed new light on this bilateral trend by separating it into two hemisphere-specific effects. The strongest effect was an underactivation of the right primary auditory cortex during state stuttering. A weaker underactivation of the left primary auditory cortex was observed for trait stuttering. Because stuttered speech for some individuals includes frequent blocking, it is unclear whether auditory underactivation in state stuttering reflects abnormal auditory processing or simply reduced auditory self-stimulation resulting from the cessation of speech. Trait stuttering, by contrast, presents no such uncertainty as it only includes fluent speech that is matched in acoustic content to the speech of PWNS in the same analyses.

Auditory areas are connected with the vocal motor system through a projection to the inferior frontal gyrus via the arcuate fasciculus (Rilling et al., 2008). This anatomical pathway is reduced bilaterally in PWS (Chang et al., 2008; Connally et al., 2014), which might be suggestive of a feed-forward deficiency. Given that the right IFG was not shown to be overactivated in state stuttering but the right larynx motor cortex was, this creates problems for a simple feed-forward pathway from auditory cortex via IFG to motor cortex. The fact that acoustic stimuli such as white noise can greatly enhance fluency in PWS suggests that the auditory system does indeed have an important feed-forward influence on the motor system. Regarding feedback, magnetoencephelography reveals that PWS have intact speech-induced suppression of

auditory responses, although with somewhat more rapid auditory responses in the right hemisphere (Beal et al., 2010). However, the fact that state stuttering was associated both with overactivation of the larynx motor cortex and with underactivation of the auditory cortex in the right hemisphere might suggest a causal connection between these two results through feedback suppression, although, as mentioned above, part of the underactivation of auditory areas in state stuttering may be due to reduced self-stimulation due to stuttering itself. Further work is needed to clarify the audio-motor relationship in stuttering, most especially disambiguating feed-forward vs. feedback contributions to stuttering. Importantly, the fact that acoustic stimuli alone can reduce the symptoms of stuttering suggests that there must be neural mechanisms for harnessing a motor system that is intrinsically overactive or discoordinated in PWS. One contributor to such a mechanism might be IFG pars orbitalis.

5.5.3 IFG pars orbitalis

The meta-analysis showed that the right IFG pars orbitalis (BA 47) was more likely to overactivate in PWS than in PWNS. A nearby, although non-overlapping, overactivation was also observed in the IFG pars orbitalis during stuttered speech compared to fluent speech in PWS. However, several studies have shown activation in this region to be negatively correlated with (trait) stuttering severity (Preibisch et al., 2003; Kell et al., 2009); the lesser the stuttering severity, the greater the activation in this region. Furthermore, activation of the IFG pars orbitalis is negatively correlated with the (state) quantity of stuttering in individual speech samples (Braun et al., 1997). PWS

engage the IFG pars orbitalis when stuttering is relatively light and fail to activate it during strong bouts of stuttering. These findings have led some researchers to speculate that activation of the IFG pars orbitalis may compensate for a dysfunction in adjacent Broca's area (Kell et al., 2009). Indeed, we found that the right homologue of Broca's area in the frontal operculum was overactive in trait stuttering, an effect seen in the Brown et al. (2005) meta-analysis as well. Neumann et al. (2005) and Kell et al. (2009), in comparing the profiles of brain activation in PWS before and after speech therapy, found an increase in right IFG pars orbitalis activity after successful therapy. This region, therefore, might provide a suppressing mechanism to the vocal motor system in Broca's area and the primary motor cortex in PWS. Such a mechanism might shed light onto the mystery of how stuttering can be ameliorated instantaneously but transiently by a diverse array of seemingly unrelated environmental and contextual factors. Further research is required to elucidate the role of the IFG pars orbitalis as a protective factor against stuttering.

5.5.4 Lip motor cortex

We observed an increase in activation in a region of the motor cortex during stuttered speech (x = 54, y = -14, z = 34) near the somatotopic lip area (x = 57, y = -10, z = 32; Brown et al., 2008) directly lateral to the larynx area. Electromyographical studies have demonstrated that, even during fluent speech, PWS are slow to articulate labial consonants (Zimmermann, 1980). The latency between initiating articulation and achieving maximal displacement of both the lower lip (van Lieshout et al., 1993) and upper lip (van Lieshout et

al., 1996) is longer in the fluent speech of PWS than PWNS. During a stutter, a slow tremor is sometimes observed in the bottom lip, although antagonistic elevator and depressor muscles still activated reciprocally as they do during fluent speech (McClean & Goldsmith, 1984).

5.5.5 Larynx

Given the general differences in activation profile observed between trait and state stuttering in our two meta-analyses, an important commonality between the two analyses was the larynx motor cortex, the principal vocal center of the human brain (Brown et al., 2008). As shown in Fig. 2, state stuttering was associated with overactivation in the right hemisphere and trait stuttering with underactivation in the homologous region of the left hemisphere.

Before discussing a laryngeal contribution to stuttering, it is important to note that motor theories of stuttering have presented evidence for disturbances at numerous levels in the speech production system, including motor timing (Alm, 2004), planning (Postma & Kolk, 1993; Howell, 2004) and articulatory control (Namasivayam & van Lieshout, 2011). In discussing a laryngeal mechanism for stuttering, we are in no way trying to discount other mechanisms or to prioritise laryngeal mechanisms over them. We are simply trying to interpret the ALE results in the most direct manner possible.

Research into the role of the larynx in stuttering declined after the 1980s when studies suggested that (i) stuttering is reduced but not eliminated when speaking in the absence of phonation, as in whispering (Perkins et al., 1976; Bruce & Adams, 1978), (ii) paralysing the larynx by injecting botulinum toxin

yields only a short-term reduction in stuttering (Ludlow, 1990; Brin et al., 1994), although the timeframe is typical of botulinum toxin treatments of neuromuscular disorders (Blitzer & Sulica, 2001), and (iii) stuttering may still occur when the larynx is excised (Tuck, 1979; although see Wingate, 1981). Interestingly, larynx excision among PWNS can result in adult-onset stuttering (Freeman & Rosenfield, 1982; Rosenfield & Freeman, 1983). Together, these findings suggest that laryngeal dysfunction may be a sufficient, but not necessary, cause of stuttering. Indeed, "prolonged speech" is a prominent stuttering therapy in which patterns of phonation are shaped to facilitate fluency (Goldiamond, 1965; Ingham, 1987). Similarly, speech conditions such as choral, rhythmic or whispered speech may induce fluency because they reduce alternations between voiced and unvoiced speech sounds (Ingham et al., 2012a).

The larynx is a complex structure with many interdependent muscles whose coordinated operations are critical to both airway protection and speech. However, laryngeal function relevant to vocalisation involves two major dimensions of muscle control − on the one hand, adduction vs. abduction of the vocal folds and, on the other, tensing vs. relaxing. While the latter is intimately associated with vocalisation, the former occurs during other processes as well, most notably during respiration and swallowing. In analysing the cortical control of these muscles, Brown et al. (2008) had subjects perform both non-vocal (glottal stops) and vocal (phonation) laryngeal tasks, and showed that the same part of the motor cortex was activated by both types of tasks, leading to a characterization of a multi-functional larynx motor cortex (for a related set of

observations, see Loucks et al., 2007 and Belyk & Brown, 2014). Two distinct regions of activation were found, namely a ventromedial peak in the primary motor cortex (BA 4) and a dorsolateral peak in the premotor cortex (BA 6). The two meta-analyses performed in the current study specifically implicated the ventromedial primary motor peak in stuttering, consistent with the results of Brown et al. (2005).

The trait underactivation of the left larynx motor cortex among PWS may be associated with trait-related deficits in the operation of the larynx. PWS are slower to initiate phonation than PWNS (Adams & Hayden, 1976). Precise timing of voicing onset is important for conveying phonetic distinctions between particular consonants (Lisker & Abramson, 1966), and voice onset times are slower (Hillman & Gilber, 1977; Zimmermann, 1980) and more variable (Jäncke, 1994) in the speech of PWS than in the speech of PWNS. In addition, problems in initiating phonation represent a key deficit among PWS, one that may trigger instances of stuttering. PWS are more likely to stutter at the beginning of an utterance or after a pause (Wall et al., 1981) as well as when speech requires alternations between voiced and unvoiced sounds, compared to when speech is voiced continuously (Adams & Reis, 1971) such as during singing. Rehearsing spoken material increases fluency, but only if the rehearsal is out loud and voiced (Brenner et al., 1972), which suggests that rehearsal aids in phonation rather than articulation. These observations in no way exclude the possibility that PWS have disordered control of the articulators. Indeed, stuttering is associated with tremor in the jaw (Platt & Iwo, 1973) and lips (McClean & Goldsmith, 1984).

The state overactivation of the right larynx area during incidences of stuttering may be related to abnormal behaviour of the laryngeal muscles during stuttered speech. During part-word repetitions, the vocal folds are abducted more than during fluent speech (Conture et al., 1977). Electromyography reveals that, during stuttered speech, the adductor and abductor muscles are overactivated and fail to coordinate as an antagonistic pair, as they do during fluent speech (Freeman & Ushijima, 1978). No abnormalities are observed in the muscles controlling laryngeal tension (Smith et al., 1996). Laryngeal blocking, which is a common component of stuttering, might be related to the simultaneous contraction of the adductor and abductor muscles, resulting in high muscle tension but little movement. This specificity of physiological abnormalities to the adductor and abducter muscles is also consistent with the difficulty that PWS have in initiating phonation. Failure to initiate phonation may lead to repetitions as the speaker attempts to initiate a syllable repeatedly. Further research is required to determine whether there is a causal relationship between overactivation of the right larynx motor cortex and abnormal laryngeal-muscle physiology.

Unlike peripheral effectors, such as the hand, that can be controlled independently of the contralateral limb, a midline structure like the larynx requires symmetric and simultaneous control of the two vocal folds. For example, for phonation to occur, the vocal folds must be adducted by simultaneous bilateral contraction of the lateral cricoarytenoid muscles, the oblique arytenoid muscles and/or the unpaired transverse arytenoid muscle. Similarly, cessation of phonation requires abduction of the vocal folds via simulatenous bilateral

contraction of the posterior cricoarytenoid muscles. Indeed, asymmetric operation of the laryngeal muscles is indicative of speech motor disorders, such as unilateral upper motor neuron dysarthria (Duffy, 2005). Well-coordinated activation of the upper motor neurons of the laryngeal system is all the more critical given the intrinsically asymmetric nature of the lower motor neurons. The left recurrent laryngeal nerve, which innervates both the abductor and adductor muscles, is twice the length of the right nerve (Prades et al., 2012).

The reciproal activation−inhibition pattern in the larynx motor cortex seen in the two hemispheres might result from a process of interhemispheric inhibition, whereby a cortical region in one hemisphere inhibits its contralateral homologue by means of callosal projections. Neef et al. (2011b) found aberrant intracortical inhibition in the tongue motor cortex of PWS and our results may be indicative of a similar phenomenon in larynx motor cortex. The left larynx motor cortex is underactivated in trait stuttering which may in turn disinhibit the right larynx motor cortex resulting in overactivation. While the meta-analysis results cannot verify whether the two hemispheres are coupled in this way, further experiments could do so.

Neuroanatomical abnormalities among PWS further implicate the larynx motor cortex and its anatomical connections. Studies using diffusion tensor imaging have found reduced fractional anisotropy in the white matter adjacent to the larynx motor cortex bilaterally (Watkins et al., 2008), indicating either reduced myelination or reduced coherence of the white matter tracts. The absence of tractography data makes it unclear whether the affected fibres are derived from corticocortical connections, the descending corticobulbar tract, or

some combination of the two. However, reduced fractional anisotropy among PWS has also been reported in both the motor cortical component of the corpus callosum (Cykowski et al., 2010; Connally et al., 2014) and the left corticobulbar-corticospinal tract (Chang et al., 2008; Connally et al., 2014), suggesting that both tracts may be affected.

5.6 Conclusions

We report an updated meta-analysis of neuroimaging studies of developmental stuttering, one that adds a new distinction between stuttering as a stable trait and stuttering as a episodic behavioural state. The results of these two analyses were remarkably divergent. Trait stuttering was characterised by the well-known rightward shift in lateralization for language and speech areas. State stuttering revealed a more diverse pattern. The larynx and lip motor cortex linked the two analyses. State stuttering was associated with overactivation in the right hemisphere lip and larynx motor cortex. Trait stuttering was associated with overactivation of lip motor cortex in the right hemisphere but underactivation of larynx motor cortex in the left hemisphere. These results suggest a potential lack of coordination in the cortical control of muscles relevant to speech.

5.7 Acknowledgements

This work was funded by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to S.B. We are grateful to Roger Ingham

for critical comments on a previous version of the manuscript.

5.8 References

Adams, M.R. & Hayden, P. (1976) The ability of stutterers and nonstutterers to initial and terminate phonation during production of an isolated vowel. *J. Speech Hear. Res., 19*, 290−296.

Adams, M.R. & Reis, R. (1971) Influence of the onset of phonation on the frequency of stuttering. *J. Speech Hear. Res., 14*, 639−644.

Akkal, D., Dum, R.P. & Strick, P.L. (2007) Supplementary motor area and presupplementary motor area: targets of basal ganglia and cerebellar output. *J. Neurosci., 27*, 10659−10673.

Alm, P.A. (2004) Stuttering and the basal ganglia circuits: a critical review of possible relations. *J. Commun. Disord., 37*, 325−369

Beal, D., Cheyne, D., Gracco, V., Quraan, M.A., Taylor, M.J. & De Nil, M.F. (2010) Auditory evoked fields to vocalization during passive listening and active generation in adults who stutter. *NeuroImage, 52*, 1645−1653.

Belyk, M. & Brown, S. (2014) Somatotopy of the extrinsic laryngeal muscles in sensorimotor cortex. *Behav. Brain Res., 270*, 364−371.

Blitzer, A. & Sulica, L. (2001) Botulinum toxin: basic science and clinical uses in otolaryngology. *Laryngoscope, 111*, 218−226.

Braun, A.R., Varga, M., Stager, S., Schulz, G., Selbie, S., Maisog, J.M., Carson, R.E. & Ludlow, C.L. (1997) Altered patterns of cerebral activity during speech and language production in developmental stuttering: an H215O positron emission tomography study. *Brain, 120*, 761−784.

Brenner, N.C., Perkins, W.H. & Sodenberg, G.A. (1972) The effect of rehearsal on frequency of stuttering. *J. Speech Hear. Res., 15*, 483−486.

Brin, M., Stewart, C., Blitzer, A. & Diamond, B. (1994) Laryngeal botulinum toxin injections for disabling stuttering in adults. *Neurology, 44*, 2262−2266.

Brocklehurst, P.H., Lickley, R.J. & Corley, M. (2013) Revisiting Bloodstein's anticipatory struggle hypothesis from a psycholinguistic perspective: a variable release threshold hypothesis of stuttering. *J. Commun. Disord., 46*, 217−237.

Brown, S., Ingham, R.J., Ingham, J.C., Laird, A.R. & Fox, P.T. (2005) Stutered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. *Hum. Brain Mapp., 25*, 105−117.

Brown, S., Ngan, E. & Liotti, M. (2008) A larynx area in the human motor cortex. *Cereb. Cortex, 18*, 837−845.

Bruce, M. & Adams, M. (1978) Effects of two types of motor practice on stuttering adaptation. *J. Speech Hear. Res., 21*, 421−428.

Burghaus, L., Hilker, R., Thiel, A., Galldiks, N., Lehnhardt, F.G., Zaro-Weber, O., Strum, V. & Heiss, W.-D. (2006) Deep brain stimulation of the subthalamic nucleus reversibly deteriorates stuttering in advanced Parkinson's disease. *J. Neural Transm., 113*, 625−631.

Chang, S.-E., Erickson, K.I., Ambrose, N.G., Hasegawa-Johnson, M.A. & Ludlow, C.L. (2008) Brain anatomy differences in childhood stuttering. *NeuroImage, 39*, 1333−1344.

Chang, S.-E., Kenney, M.K., Loucks, T.M.J. & Ludlow, C.L. (2009) Brain activation abnormalities during speech and non-speech in stuttering speakers. *NeuroImage, 46*, 201−212.

Commodore, R.W. & Cooper, E.B. (1978) Communicative stress and stuttering frequency during normal, whispered, and articulation-without phonation speech modes. *J. Fluency Disord., 3*,1−12.

Connally, E.L., Ward, D., Howell, P. & Watkins, K.E. (2014) Disrupted white matter in language and motor tracts in developmental stuttering. *Brain Lang., 131*,25−35.

Conture, E.G., McCall, G.N. & Brewer, D.W. (1977) Laryngeal behaviour during stuttering. *J. Speech Hear. Res., 20*, 661−668.

Cykowski, M.D., Fox, P.T., Ingham, R.J., Ingham, J.C. & Robin, D.A. (2010)

A study of the reproducibility and etiology of diffusion anisotropy differences in developmental stuttering: a potential role for impaired myelination. *NeuroImage, 52*, 1495−1504.

Davidow, J.H., Bothe, A.K., Andreatta, R.D. & Ye, J. (2009) Measurement of phonated intervals during four fluency-inducing conditions. *J. Speech Lang. Hear. R., 52*, 188−205.

De Nil, L.F., Kroll, R.M., Kapur, S. & Houle, S. (2000) A positron emission tomography of silent and oral word reading in stuttering and nonstuttering adults. *J. Speech Lang. Hear. R., 43*, 1038−1053.

De Nil, L.F., Kroll, R., Lafaille, S.J. & Houle, S. (2003) A positron emission tomography study of short and long-term treatment effects on functional brain activation in adults who stutter. *J. Fluency Disord., 28*, 357−380.

De Nil, L.F., Beal, D.S., Lafaille, S.J., Kroll, R.M., Crawley, A.P. & Gracco, V.L. (2008) The effects of simulated stuttering and prolonged speech on the neural activation patterns of stuttering and nonstuttering adults. *Brain Lang., 107*, 114−123.

Duffy, J.R. (2005) Motor Speech Disorders, 2nd Edn. Elsevier, USA. Dum, R.P. & Strick, P.L. (1991) The origin of corticospinal projections from the premotor areas in the frontal lobe. *J. Neurosci., 11*, 667−689.

Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K. & Fox, P.T.L. (2009) Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Hum. Brain Mapp., 30*, 2907− 2926.

Forster, D.C. & Webster, W.G. (2001) Speech-motor control and interhemispheric relations in recovered and persistent stuttering. *Dev. Neuropsychol., 19*, 125−145.

Fox, P., Ingham, R., Ingham, J. & Hirsch, T. (1996) A PET study of the neural systems of stuttering. *Nature, 382*, 158−162.

Fox, P.T., Ingham, R.J., Ingham, J.C., Zamarripa, F., Xiong, J.-H. & Lancaster, J.L. (2000) Brain correlates of stuttering and syllable production. A PET performance-correlation analysis. *Brain, 123*, 1985−2004.

Freeman, F.J. & Rosenfield, D.B. (1982) Source in dysfluency. *J. Fluency Disord., 7*, 295−296.

Freeman, F. & Ushijima, T. (1978) Laryngeal muscle activity during stuttering. *J. Speech Hear. Res., 21*, 538−562.

Garber, S. & Martin, R.R. (1977) Effects of noise and increased vocal intensity on stuttering. *J. Speech Hear. Res., 20*, 233−240.

Giraud, A.-L., Neumann, K., Bachoud-Levi, A.-C., von Gudenberg, A.W., Euler, H.A., Lanfermann, H. & Preibisch, C. (2008) Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. *Brain Lang., 104*, 190−199.

Goldiamond, I. (1965) Stuttering and fluency as manipulable operant response classes. In Krasner, L. & Ullman, L. (Eds), *Research in Behaviour Modification.* Holt, Rinehart & Winston, New York, pp. 106−156.

Hillman, R.E. & Gilber, H.R. (1977) Voice onset time for voiceless stop consonants in the fluent reading of stutterers and nonstutterers. *J. Acoust. Soc. Am., 61*, 610−611.

Howell, P. (2004) Assessment of some contemporary theories of stuttering that apply to spontaneous speech. Contemp. *Issues Commun. Sci. Disord., 31*, 122−139.

Howell, P., Jiang, J., Peng, D. & Lu, C. (2012) Neural control of rising and falling tones in Mandarin speakers who stutter. *Brain Lang., 123*, 211−221.

Inase, M., Tokuno, H., Nambu, A., Akazawa, T. & Takada, M. (1999) Corticostriatal and corticosubthalamic input zones from the presupplementary motor area in the macaque monkey: comparison with the input zones from the supplementary motor area. *Brain Res., 833*, 191−201.

Ingham, R. (1987) Residential prolonged speech stutering therapy manual. Department of Speech and Hearing Sciences, University of California, Santa Barbara, CA.

Ingham, R.J., Fox, P.T., Ingham, J.C., Xiong, J., Zamarripa, F., Hardies, L.J.

& Lancaster, J.L. (2004) Brain correlates of stuttering and syllable production: gender comparison and replication. J. Speech Lang. Hear. R., 47, 321−341.

Ingham, R., Bothe, A.K., Wang, Y., Purkhiser, K. & New, A. (2012a) Phonation interval modification and speech performance quality during fluency-inducing conditions by adults who stutter. *J. Commun. Disord., 45*, 198−211.

Ingham, R.J., Grafton, S.T., Bothe, A.K. & Ingham, J.C. (2012b) Brain activity in adults who stutter: similarities across speaking tasks and correlations with stuttering frequency and speaking rate. *Brain Lang., 122*, 11−24.

Jäncke, L. (1994) Variability and duration of voice onset time and phonation in stuttering and nonstuttering adults. *J. Fluency Disord., 19*,21−37.

Jiang, J., Lu, C., Peng, D., Zhu, C. & Howell, P. (2012) Classification of types of stuttering symptoms based on brain activity. *PLoS One, 7*, e39747.

Johnson, W. & Rosen, L. (1937) Studies in the psychology of stuttering: VII: effects if certain changes in speech pattern upon frequency of stuttering. *J. Speech Hear. Res., 2*, 105−110.

Kell, C.A., Neumann, K., von Kriegstein, K., Posenenske, C., von Gudenberg, A.W., Euler, H. & Giraud, A.-L. (2009) How the brain repairs stuttering. *Brain, 132*, 2747−2760.

Klouda, G.V. & Cooper, W.E. (1988) Contrastive stress, intonation, and stuttering frequency. *Lang. Speech, 31*,3−20.

Kraft, S.J. & Yairi, E. (2012) Genetic bases of stuttering: the state of the art, 2011. *Folia Phoniatr. Logo., 64*,34−47.

van Lieshout, P.H., Peters, H.F., Starkweather, C.W. & Hulstijn, W. (1993) Physiological differences between stutterers and nonstutterers in perceptually fluent speech: EMG amplitude and duration. *J. Speech Hear. Res., 36*, 55−63.

van Lieshout, P.H., Hulstihn, W. & Peters, H.F.M. (1996) From planning to articulation: testing the motor plan assembly hypothesis. *J. Speech Lang. Hear. R., 39*, 546−564.

Lisker, L. & Abramson, A.S. (1966) Some effects of context on voice onset

time in English stops. *Lang. Speech, 10*,1−28.

Loucks, T.M.J., Poletto, C.J., Simonyan, K., Reynolds, C.L. & Ludlow, C.L. (2007) Human brain activation during phonation and exhalation: common volitional control for two upper airway functions. *NeuroImage, 36*, 131− 143.

Loucks, T., Kraft, S.J., Choo, A.L., Sharma, H. & Ambrose, N.G. (2011) Functional brain activation differences in stuttering identified with a rapid fMRI sequence. *J. Fluency Disord., 36*, 302−307.

Lu, C., Chen, C., Ning, N., Ding, G., Guo, T., Peng, D., Yang, Y., Li, K. & Lin, C. (2010) The neural substrates for atypical planning and execution of word production in stuttering. *Exp. Neurol., 221*, 146−156.

Ludlow, C.L. (1990) Treatment of speech and voice disorders with botulinum toxin. *JAMA, 264*, 2671−26715.

Luppino, G., Matelli, M. & Camarda, R. (1993) Corticocortical connections of area F3 in the macaque monkey. *J. Comp. Neurol., 140*, 114−140.

Max, L., Guenther, F.H., Gracco, V.L., Ghosh, S.S. & Wallace, M.E. (2004) Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: a theoretical model of stuttering. Contemp. *Issues Commun. Sci. Disord., 31*, 105−122.

McClean, M. & Goldsmith, H. (1984) Lower-lip EMG and displacement during bilabial disfluencies in adult stutterers. *J. Speech Hear. Res., 27*, 342−349.

Namasivayam, A. & van Lieshout, P. (2011) Speech motor skill and stuttering. *J. Motor Behav., 43*,37−41.

Neef, N.E., Jung, K., Rothkegel, H., Pollock, B., von Gudenberg, A.W., Paulus, W. & Sommer, M. (2011a) Right-shift for non-speech motor processing in adults who stutter. *Cortex, 47*, 945−954.

Neef, N.E., Paulus, W., Neef, A., von Gudenberg, A.W. & Sommer, M. (2011b) Reduced intracortical inhibition and facilitation in the primary motor tongue representation of adults who stutter. *Clin. Neurophysiol., 122*, 1802−1811.

Neumann, K., Euler, H.A., von Gudenberg, A.W., Giraud, A.-L., Lanfermann,

H., Gall, V. & Preibisch, C. (2003) The nature and treatment of stuttering as revealed by fMRI. *J. Fluency Disord., 28*, 381−410.

Neumann, K., Preibisch, C., Euler, H.A., von Gudenberg, A.W., Lanfermann, H., Gall, V. & Giraud, A.-L. (2005) Cortical plasticity associated with stuttering therapy. *J. Fluency Disord., 30*,23−39.

Nippold, M. (2012) Stuttering and language ability in children: questioning the connection. *Am. J. Speech-Lang. Pat., 21*, 183−197.

Ntourou, K., Conture, E.G. & Lipsey, M.W. (2011) Language abilities of children who stutter: a meta-analytic review. *Am. J. Speech-Lang. Pat., 20*, 163−180.

den Ouden, D.-B., Montgomery, A. & Adams, C. (2013) Simulating the neural correlates of stuttering. *Neurocase, 20*,37−41.

Perkins, W., Rudas, J., Johnson, L. & Bell, J. (1976) Stuttering: discoordination of phonation with articulation and respiration. *J. Speech Hear. Res., 19*, 509−522.

Platt, J. & Iwo, A.I. (1973) Jaw tremor during stuttering block: an electromyographic study. *J. Commun. Disord., 6*, 102−109.

Postma, A. & Kolk, H. (1993) The covert repair hypothesis: prearticulatory repair processes in normal and stuttered disfluencies. *J. Speech Hear. Res., 36*, 472−487.

Prades, J.M., Dubois, M.D., Dumollard, J.M., Tordella, L., Rigail, J., Timoshenko, A.P. & Poec'h, M. (2012) Morphological and functional asymmetry of the human recurrent laryngeal nerve. *Surg. Radiol. Anat., 34*, 903−908.

Preibisch, C., Neumann, K., Raab, P., Euler, H.A., von Gudenberg, A.W., Lanfermann, H. & Giraud, A.-L. (2003) Evidence for compensation for stuttering by the right frontal operculum. *NeuroImage, 20*, 1356−1364.

Rappaport, B. & Bloodstein, O. (1971) The role of random blackout cues in the distribution of moments of stuttering. *J. Speech Hear. Res., 14*, 874− 879.

Rilling, J.K., Glasser, M.F., Preuss, T.M., Ma, X., Zhao, T., Hu, X. & Beh-

rens, T.E.J. (2008) Evolution of the arcuate fasciculus. *Nat. Neurosci., 11*, 426−428.

Rosenfield, D.B. & Freeman, F.J. (1983) Stuttering onset after laryngectomy. *J. Fluency Disord., 8*, 265−268.

Sakai, N., Masuda, S., Shimotomai, T. & Mori, K. (2009) Brain activation in adults who stutter under delayed auditory feedback: an fMRI study. Int. *J. Speech Lang. Pathol., 11*,2−11.

Smith, A., Denny, M., Shaffer, L.A., Kelly, E.M. & Hirano, M. (1996) Activity of intrinsic laryngeal muscles in fluent and disfluent speech. *J. Speech Hear. Res., 39*, 329−348.

Stuart, A., Kalinowski, J. & Rastatter, M.P. (1997) Effect of monaural and binaural altered auditory feedback on stuttering frequency. *J. Acoust. Soc. Am., 101*, 3806−3809.

Talairach, J. & Tournoux, P. (1988) Co-planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cere- bral Imaging. Georg Thieme Verlag, New York.

Toscher, M.M. & Rupp, R.R. (1978) A study of the central auditory processes in stutterers using the Synthetic Sentence Identification (SSI) test battery. *J. Speech Hear. Res., 21*, 779−792.

Toyomura, A., Fujii, T. & Kuriki, S. (2011) Effect of external auditory pacing on the neural activity of stuttering speakers. *NeuroImage, 57*, 1507−1516.

Travis, L.E. (1978) The cerebral dominance theory of stuttering: 1931−1978. *J. Speech Hear. Res., 43*, 279−281.

Tuck, A.E. (1979) An alaryngeal stutterer: a case history. *J. Fluency Disord., 4*, 239−243.

Turkeltaub, P., Eden, G.F., Jones, K.M. & Zeffiro, T.A. (2002) Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage, 16*, 765−780.

Turkeltaub, P.E., Eickhoff, S.B., Laird, A.R., Fox, M., Wiener, M. & Fox, P.

(2011) Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Hum. Brain Mapp., 33*,1−13.

Wall, M.J., Starkweather, C.W. & Harris, K.S. (1981) The influence of voicing adjustments on the location of stuttering in the spontaneous speech of young child stutterers. *J. Fluency Disord., 6*, 299−310.

Watkins, K.E., Smith, S.M., Davis, S. & Howell, P. (2008) Structural and functional abnormalities of the motor system in developmental stuttering. *Brain, 131*,50−59.

Wingate, M.E. (1964) A standard definition of stuttering. *J. Speech Hear. Disord., 29*, 484−489.

Wingate, E. (1981) Questionnaire study of laryngectomee stutterers. *J. Fluency Disord., 6*, 273−281.

Wu, J., Maguire, G., Riley, G. & Fallon, J. (1995) A position emission tomography [18F]deoxyglucose study of developmental stuttering. *NeuroReport, 6*, 501−505.

Wymbs, N.F., Ingham, R.J., Ingham, J.C., Paolini, K.E. & Grafton, S.T. (2013) Individual differences in neural regions functionally related to real and imagined stuttering. *Brain Lang., 124*, 153−164.

Yairi, E. (2007) Subtyping stuttering I: a review. *J. Fluency Disord., 32*, 165−196.

Zimmermann, G. (1980) Articulatory dynamics of fluent utterances of stutterers and nonstutterers. *J. Speech Lang. Hear. R., 23*,95−107.

Chapter 6

General discussion

Michel Belyk

The neural control of the larynx and the voice in humans is an understudied topic. This is perhaps surprising considering the prevalence of research related to the neuroscience of speech and the critical role that the larynx plays as the principal sound source for vocalization. In light of the paucity of research on the human vocal-motor system, this dissertation described a series of brain-imaging experiments and meta-analyses that elucidate the organization of this system in humans.

Chapter two characterized brain areas specialized for making explicit judgments about the affective content of heard vocalizations compared to other aspects of speech. This meta-analysis found that the IFG pars orbitalis was a key brain area for perceiving affective vocalizations. Since prefrontal brain areas that are involved in perceiving actions are often involved in producing the same actions (Cross, Hamilton, & Grafton, 2006; Watkins & Paus, 2004), this

chapter hypothesized that the IFG pars orbitalis may participate in planning affective vocalizations. This hypothesis was falsified by experiments reported in chapter three.

Chapter three tested the predictions of non-human primate models of the vocal-motor system. While these species are appealing as models, owing to their taxonomic proximity to humans, they are a poor match to the human vocal phenotype since they lack the capacity for vocal production learning (Hayes & Hayes, 1951). Chapter three failed to support predictions about the human vocal-motor system derived from the vocal system of monkeys. In particular, the human vocal-motor system does not appear to be divided into separate cingulate and primary motor vocal pathways for affective and speech-like vocalizations, respectively, but rather consists of a single network that controls the vocal apparatus.

Chapter four supported the songbird vocal system as an alternative model of the vocal-motor system's capacity for vocal production learning. In songbirds, a specialized basal ganglia pathway is critical for vocal learning via imitation, and it has previously been hypothesized that this pathway has an analogue in the human brain (Jarvis, 2007). Chapter four demonstrated that the same system is preferentially engaged in the human brain during vocal imitation although further research is required to link deficits in imitative abilities to dysfunction in this pathway.

Chapter five reported a meta-analysis of brain-imaging studies of persistent developmental stuttering and further subdivided these findings into abnormal activation related to the trait of being a person who stutters versus the state

of exhibiting a stutter. The right dorsal larynx motor area, among other brain regions, was overactivated in people who stutter while the same area in the left hemisphere was underactivated during a stuttering event. This suggests a deviation from the usual bilateral symmetry of activation. The Rolandic operculum was also abnormally activated in people who stutter, although the location of this result was slightly removed from the expected position of the ventral larynx motor area. The mechanisms through which aberrant cortical motor activity translates to the act of stuttering remains unknown and will require a greater knowledge of the human vocal-motor system to link variation in the structure and function of the this system to perturbations in speech.

The following sections outline fundamental gaps in our knowledge of the vocal-motor system. Further research addressing these questions would contribute greatly to our understanding of the neural mechanisms controlling the larynx and how it has changed over the course of primate evolution.

6.1 Evolution of the larynx motor cortex

Although it is clear that the laryngeal motor cortex has been considerably altered over the course of primate evolution, the progression from a single premotor region with minimal involvement in vocalization (Jürgens, 2002) to a pair of primary motor areas (Bouchard et al., 2013) that are necessary for vocalizations (Jürgens et al., 1982), remains poorly defined. An early brain imaging study that outlined the dorsal larynx motor area as distinct from surrounding orofacial motor areas (Brown et al., 2008) suggested that the larynx area may have migrated through primate evolution from its location in

ventral premotor cortex in monkeys (Jürgens, 2009) to the border of ventral primary motor cortex in non-human apes (Leyton & Sherrington, 1917) to dorsal primary motor cortex in humans. However, in light of more recent research that has outlined a second larynx-motor area in ventral primary motor cortex (Bouchard et al., 2013), models of this evolutionary progression may require revision.

Two lines of evidence suggest that the ventral − rather than dorsal − larynx area in humans may be homologous to the non-human primate larynx area. First, the greater proximity of the ventral larynx area to that of non-human primates, particularly to that of other apes (Leyton & Sherrington, 1917), suggests a neural migration that traverses a smaller distance. Second, electrical stimulation of the ventral larynx area elicits grunting sounds (Foerster, 1931) similar to those elicited in non-human apes (Leyton & Sherrington, 1917), in contrast to the speech-like vowel sounds elicited by stimulation of the dorsal larynx area (Penfield & Boldrey, 1937).

Under this hypothesis, the human dorsal larynx area may be a novel functional region with no homologue in other primates. The proximity of the dorsal larynx area to motor areas for the articulatory and respiratory musculature may facilitate short latency connections that permit efficient communication between motor regions that control the muscles of speech. However, the paucity of electrical stimulation experiments in humans and other apes as well as a dependence on natural experiments in lieu of lesion studies makes hypotheses regarding inter-species differences difficult to test. Furthermore, the evidence that is currently available comes from only a handful of species,

representing a sparse sample of the primate order. A rigorous test of such evolutionary hypotheses will require broader sampling of species across the primate order and the application of statistical methods that control for phylogenetic relationships.

6.2 One larynx, but two larynx motor areas

While the division of the larynx motor cortex into two non-contiguous regions suggests a division of function, brain imaging studies have so far failed to dissociate the functions of these areas. Both larynx motor areas are active during phonation (Bouchard et al., 2013; Brown et al., 2009, 2008; Grabski et al., 2012; Olthoff, Baudewig, Kruse, & Dechent, 2008) and forced expiration (Loucks et al., 2007; Ramsay et al., 1993). Few studies have examined the neural correlates of glottal stops, which involve a forceful adduction of the vocal folds by engaging the lateral cricoarytenoid and interarytenoid muscles. Brown et al. (2008), which is the only study that has examined this movement in isolation, observed activation of both larynx areas for glottal stops, but only the dorsal larynx area for phonation. Coughing includes similar movements and activates both larynx motor areas (Mazzone, Cole, Ando, Egan, & Farrell, 2011; Simonyan et al., 2007). Glottal stops embedded within the consonant-vowel-consonant syllable /i?i/ also produces similar activations, spanning both larynx motor areas (Simonyan, Ostuni, Ludlow, & Horwitz, 2009). Vertical movement of the larynx in either direction engages the bulk of the orofacial motor cortex, including both dorsal and ventral larynx motor areas (Belyk & Brown, 2014). This literature suggests that the dorsal and ventral larynx areas

may both be involved in controlling a suite of speech motor functions, including phonation, expiration, adduction/abduction of the vocal folds, and vertical positioning of the larynx. However, few studies have examined more than one of these functions in the same set of participants, making co-localization of function difficult to assess.

One species difference that is frequently cited as a key adaptation in the human brain is the emergence of direct projections from larynx motor cortex to lower motor neurons in the brainstem (Fitch, 2010). While evidence for the absence of this direct pathway in monkeys (Jürgens & Ehrenreich, 2007; Kristina Simonyan & Jürgens, 2003) and the sparseness of this pathway in non-human apes (Kuypers, 1958b) is derived from well controlled studies, this pathway has been assessed in humans only from natural experiments. Two natural experiments have traced deteriorating fibers originating in damaged cortex directly to the nucleus ambiguus in humans (Iwatsubo et al., 1990; Kuypers, 1958a). However, in both of these studies, cortical lesions extended to both dorsal and ventral larynx areas. It is therefore not known whether cortical projections to the nucleus ambiguus originate in the dorsal, ventral or both larynx motor areas.

6.3 The human cingulate vocal area

Based on homology with monkeys, the human anterior cingulate cortex has previously been hypothesized to contain a region specialized for the volitional

production of affective vocalization (Ackermann, Hage, & Ziegler, 2014; Myers, 1976; Owren, Amoss, & Rendall, 2011). Indeed in monkeys, this region projects to the PAG, which is a critical structure for organizing affective vocal-motor patterns (Jürgens & Pratt, 1979a, 1979b) and lesion studies have demonstrated that the ACC is necessary for producing affective vocalizations outside of contexts that would normally elicit those vocalizations (Aitken, 1981; Sutton et al., 1974, 1981). While such findings suggest that the ACC is indeed involved in the volitional production of affective vocalizations, they say little about specialization for that function, particularly when generalized to humans, who differ markedly from other primates in their vocal capabilities.

The vocal repertoire of non-human primates consists entirely of species-specific affective vocalization. Humans, however, have a broader vocal repertoire that is acquired via vocal production learning. Although brain imaging studies in humans have observed that the ACC is indeed activated by producing affective vocalizations (Barrett et al., 2004; Wattendorf et al., 2013), it is also activated when producing learned vocal patterns that are unrelated to affect (Belyk & Brown, In Press). Hence, the vocal anterior cingulate cortex appears to be more broadly related to vocalization than previously supposed.

The anterior cingulate sulcus contains multiple motor areas along its dorsal and ventral banks (Dum & Strick, 1991). Each of these subdivisions is organized somatotopically, with the feet represented caudally and the face represented rostrally (Amiez & Petrides, 2014). The more anterior of these cingulate motor areas exists in the approximate location of the cingulate vocalization area above the genu of the corpus callosum. This region has been

referred to variously as the rostral cingulate motor area in rhesus monkeys (Picard & Strick, 1996), the anterior rostral cingulate zone in humans (Picard & Strick, 1996), and the anterior middle cingulate cortex in both species (Vogt, Berger, & Derbyshire, 2003; Vogt, Vogt, Farber, & Bush, 2005). In contrast to its caudal counterpart, which is involved in initiating simple movements, this region may be involved in selecting actions among multiple alternatives (Mueller, Brass, Waszak, & Prinz, 2007; Shima, 1998). Hence, one possibility is that the cingulate vocalization area may be the somatotopic larynx division of a cingulate motor area that plays a role in selecting vocal patterns, affective or otherwise. This would be consistent with the established role of this region in vocal usage learning (Aitken, 1981; Sutton et al., 1974, 1981). Further research should test this hypothesis by comparing the functional properties of the vocal anterior cingulate sulcus to other somatotopic subdivisions of the same region.

6.4 Missing white matter pathways

6.4.1 Closing the audio-vocal loop

All models of the vocal system implicitly assume that the auditory and vocal systems form a loop with information flowing from auditory areas in the temporal lobe towards motor areas in the frontal lobe. For instance, the classic Wernicke-Geschwind model (Geschwind, 1970) posits that the arcuate fasciculus, which links the superior temporal gyrus with the inferior frontal gyrus, is critical for vocal imitation. Since, the primary motor cortex is a requisite

brain structure for vocalization in humans (Jürgens et al., 1982), the Wernicke-Geschwind model implicitly assumes either direct or indirect efferent projections from the inferior frontal gyrus to primary motor cortex for the execution of vocal-motor patterns. Similarly, the computational model Directions Into Velocities of Articulators treats the same pathway as part of a loop that transmits auditory feedback from self-produced vocalizations (Golfinopoulos, Tourville, & Guenther, 2010). However, anatomical connections between the IFG and the larynx motor areas are poorly described in humans.

Some evidence for functional connectivity between the IFG and the orofacial motor cortex comes from one surgical study in which electrical stimulation was applied to the IFG which then evoked responses in ventral primary motor cortex (Greenlee et al., 2004). Subsequent stimulation of those areas of motor cortex in which a response was observed produced a broad range of orofacial movements. Speech arrest and changes in vocal pitch were elicited from one such region near the ventral extent of the precentral gyrus suggestive of the ventral larynx area.

6.4.2 The corpus callosum

Inter-hemispheric fibers in division III of the corpus callosum connect the primary motor cortex in the two cerebral hemispheres (Fling, Benson, & Seidler, 2013; Hofer & Frahm, 2006). The corpus callosum contains a motor homunculus akin to that in M1, with the legs represented posteriorly and the face represented anteriorly (Wahl et al., 2007). However, no study has reported such fibers for either larynx motor area in humans. This may be due either to

the absence of such fibers or to limitations in current methodologies.

The axons of the motor corpus callosum carry primarily inhibitory signals between homologous primary motor regions that are believed to facilitate asymmetrical movement of the two sides of the body (Netz, Ziemann, & Hömberg, 1995). Unlike the limbs, which can be controlled independently of the contralateral side, the larynx is a midline structure that operates symmetrically. For phonation to occur, the vocal folds must be adducted by simultaneous bilateral contraction of the lateral cricoarytenoid muscles, the oblique interarytenoid muscles and/or the transverse interarytenoid muscle. Similarly, cessation of phonation requires abduction of the vocal folds via simultaneous bilateral contraction of the posterior cricoarytenoid muscles.

While descending efferents from the primary motor cortex for the limbs only reach lower motor neurons on the contralateral side, efferent fibers from laryngeal motor cortex reach lower motor neurons in the nucleus ambiguus bilaterally (Kuypers, 1958a). These bilateral projections may guard against asymmetric contraction of the laryngeal muscles, which is indicative of speech motor disorders, such as unilateral upper motor neuron dysarthria (Duffy, 2005).

It is not clear what role a mechanism of inter-hemispheric inhibition, such as has been described for the limbs, would play in laryngeal motor control. Further research is required to determine first whether the larynx areas in the left and right hemispheres are connected via callosal fibers and second whether these fibers are inhibitory or if other mechanisms are in place that instead promote symmetrical activation of the laryngeal muscles. Understanding the

relationship between the larynx areas of either hemisphere is critical to our understanding of the asymmetric activation of larynx motor areas observed in persistent developmental stuttering (Belyk et al., 2015).

6.4.3 Methodological challenges

Although the inferior frontal and inter-hemispheric pathways discussed above have been described in monkeys, methodological obstacles have prevented researchers from studying these pathways in humans. Tracer studies in rhesus monkeys have demonstrated that the larynx motor area in that species has strong bidirectional connections with the ipsilateral IFG and contralateral larynx area (Simonyan & Jürgens, 2005; Kristina Simonyan & Jürgens, 2002). However, similar studies in songbirds demonstrate the absence of inter-hemispheric connections for the nucleus RA, which is analogous to human larynx motor cortex (Schmidt, Ashmore, & Vu, 2004). In vivo studies in humans using Diffusion Tensor Imaging (DTI) have failed to satisfactorily describe these pathways. Simonyan et al. (2009) localized the dorsal larynx area using fMRI and attempted to trace its white matter pathways using DTI. However, projections between the IFG and primary motor cortex were observed in only a minority of participants and projections to the contralateral motor cortex were never observed. Wahl et al. (2007) was only able to detect callosal fibers for the somatotopic lip area, which is adjacent to the dorsal larynx area, in the brain of one participant out of twelve.

One possibility is that these pathways do indeed exist in humans, but that they are difficult to detect with current technology. DTI has known

limitations in modeling crossed pathways. Although methodological advances have somewhat alleviated this difficulty (Jbabdi & Johansen-berg, 2011), it remains possible that cortico-cortical projections from either larynx motor area are masked by the corticospinal tract, which passes adjacent to ventral motor cortex.

6.5 The genetics of vocal proficiency

Just as the clinical observations of Broca and Wernicke led to early neurological descriptions of brain areas related to human language and speech, the genotyping of patients with developmental speech disorders has led to the first discovery of a gene for these functions. Mutation of FOXP2 were first described in a family with a high density of members with widespread expressive and receptive speech and language deficits (Hurst, Baraitser, Auger, Graham, & Norell, 1990; Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001; Watkins, Dronkers, & Vargha-Khadem, 2002). Those affected have abnormal brain function (Liégeois, Morgan, Connelly, & Vargha-Khadem, 2011) and structure (Watkins, Vargha-Khadem, et al., 2002) in areas supporting language and speech articulation. That FOXP2 regulates the expression of a host of other genes that affect neural development (Vernes et al., 2007, 2011) may explain how mutation of this single gene has such a wide-reaching impact on the brain and correspondingly on behaviour. However, despite the broad array of phenotypical characteristics of FOXP2 mutations, they have no known association with laryngeal function. The genetic mechanisms driving the development of the human larynx motor cortex and its divergence from that found in other

primates remain unknown, although there are several promising lines of investigation.

In examining similarities in gene expression between the avian and human vocal-motor systems, Pfenning et al. (2014) observed that several genes were downregulated in the dorsal and/or ventral larynx motor cortex and their avian analog relative to surrounding tissue in humans and vocally-imitating birds, but not in non-vocal learning primates or birds. These genes are therefore candidates for the loci of adaptations underlying the neural specialization of larynx motor cortex. Notable among these candidate genes are SLIT1 and NEUROD6, both of which have established roles in neural development (Dickson, 2002; Wu et al., 2005).

Another approach towards exploring the genetics of laryngeal motor control is to search for genes that confer a risk of developing speech disorders, such as persistent developmental stuttering, that have an established laryngeal component. Persistent developmental stuttering is heritable (Ambrose & Cox, 1997; Felsenfeld et al., 2000), associated with poor coordination of certain laryngal muscles (Freeman & Ushijima, 1978), and is associated with abnormal brain function in the dorsal larynx motor cortex, among other areas (Belyk, Kraft, & Brown, 2015; Brown, Ingham, Ingham, Laird, & Fox, 2005).

Although investigations into the genetic etiology of stuttering are just beginning, polymorphisms that may be associated with persistent developmental stuttering have been identified in several candidate genes. DRD2, which is a dopamine receptor gene (Lan et al., 2009), as well as GNPTAB, GNPTG and NAGPA, which code for molecular components of a lysosomal transport

pathway (Han et al., 2014; Kang et al., 2010) are associated with stuttering. However, mechanisms that map the molecular consequences of these mutations onto the neurophenotypes of stuttering are lacking. Furthermore, given the diversity of neural abnormalities among people who stutter (Beal, Gracco, Brettschneider, Kroll, & De Nil, 2013; Chang, Erickson, Ambrose, Hasegawa-Johnson, & Ludlow, 2008; Watkins, Smith, Davis, & Howell, 2008) it is presently unclear to which components of the disorder these polymorphisms are related. Research localizing the effects of these polymorphisms to the specific neural phenotypes of this disorder may further elucidate the genetics of the human vocal-motor system.

6.6 Concluding remarks

This dissertation has described a series of human brain imaging experiments and meta-analyses seeking to elucidate the neural mechanisms controlling the organ of vocalization − namely the larynx. Though the larynx is of critical importance to speech, which is a hallmark of our species, much of our knowledge of the neural system that controls this structure had previously been derived from non-human primate models. Contrary to the predictions of these models, the vocal-motor system as a whole participates in producing vocalization with no evidence of specialization for innate and affective versus learned and arbitrary vocal patterns. However, adaptations to parts of this network, particularly in the primary motor cortex and the putamen, may underlie the human capacity for vocal production learning via vocal imitation. The abnormal operation of this system in people who stutter may provide a promising

model of disordered vocal-motor control.

6.7 References

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences, 37*, 529−604.

Adams, M.R. & Reis, R. (1971) Influence of the onset of phonation on the frequency of stuttering. *J. Speech Hear. Res., 14*, 639−644.

Aitken, P. G. (1981). Cortical control of conditioned and spontaneous vocal behavior in rhesus monkeys. *Brain and Language, 13*(1), 171−184.

Ambrose, N. G., & Cox, N. J. (1997). The genetic basis of persistence and recovery in stuttering. *Journal of Speech, Language & Hearing, 40*(3), 567−580.

Amiez, C., & Petrides, M. (2014). Neuroimaging evidence of the anatomo-functional organization of the human cingulate motor areas. *Cerebral Cortex, 24*(3), 563−578.

Ardran, G., & Kemp, F. (1952). The protection of the laryngeal airway during swallowing. *British Journal of Radiology, 25*, 406−416.

Aronov, D., Andalman, A. S., & Fee, M. S. (2008). A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science, 320*(5876), 630−634.

Arriaga, G., & Jarvis, E. D. (2013). Mouse vocal communication system: are ultrasounds learned or innate? *Brain and Language, 124*(1), 96−116.

Aziz-Zadeh, L., Sheng, T., & Gheytanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS One, 5*(1), e8759.

Banse, R., & Sherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614−636.

Barbizet, J. (1958). Yawning. *Journal of Neurology, Neurosurgery, and Psychiatry, 21*, 203−208.

Barrett, J., Pike, G. B., & Paus, T. (2004). The role of the anterior cingulate cortex in pitch variation during sad affect. *European Journal of Neuroscience, 19*(2), 458−464.

Beal, D. S., Gracco, V. L., Brettschneider, J., Kroll, R. M., & De Nil, L. F. (2013). A voxel-based morphometry (VBM) analysis of regional grey and white matter volume abnormalities within the speech production network of children who stutter. *Cortex, 49*(8), 2151−61.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40*(2), 531−539.

Belyk, M., & Brown, S. (In Press). Pitch underlies activation of the vocal system during affective vocalization. *Social Cognitive and Affective Neuroscience*.

Belyk, M., & Brown, S. (2014). Somatotopy of the extrinsic laryngeal muscles in the human sensorimotor cortex. *Behavioural Brain Research, 270*, 364−371.

Belyk, M., Kraft, S. J., & Brown, S. (2015). Stuttering as a trait or state - an ALE meta-analysis of neuroimaging studies. *The European Journal of Neuroscience, 41*(2), 275−284.

Bergman, T. J. (2013). Speech-like vocalized lip-smacking in geladas. *Current Biology, 23*(7), R268−R2689.

Blumin, J. H., Pcolinsky, D. E., & Atkins, J. P. (2004). Laryngeal findings in advanced Parkinson's disease. *Annals of Otology, Rhinology and Laryngology, 113*, 253−258.

Bottjer, S., Miesner, E., & Arnold, A. (1984). Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science, 224*(4651), 901−903.

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature, 495*(21), 327−332.

Breshears, J. D., Molinaro, A. M., & Chang, E. F. (In Press). A probabilistic map of the human ventral sensorimotor cortex using electrical stimulation. *Journal of Neurosurgery.*

Brown, S., Ingham, R. J., Ingham, J. C., Laird, A. R., & Fox, P. T. (2005). Stuttered and fluent speech production: An ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping, 25*(1), 105−117.

Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Turkeltaub, P., & Liotti, M. (2009). The somatotopy of speech: Phonation and articulation in the human motor cortex. *Brain and Cognition, 70*(1), 31−41.

Brown, S., Martinez, M. J., Hodges, D. a, Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. *Cognitive Brain Research, 20*(3), 363−375.

Brown, S., Ngan, E., & Liotti, M. (2008). A larynx area in the human motor cortex. *Cerebral Cortex, 18*(4), 837−845.

Canter, G. J. (1963). Speech characteristics of patients with Parkinson's disease: I. intensity, pitch and duration. *The Journal of Speech and Hearing Disorders, 28*, 221−229.

Chang, S.-E., Erickson, K. I., Ambrose, N. G., Hasegawa-Johnson, M. A., & Ludlow, C. L. (2008). Brain anatomy differences in childhood stuttering. *NeuroImage, 39*(3), 1333−1344.

Coudé, G., Ferrari, P. F., Rod, F., Maranesi, M., Borelli, E., Veroni, V., Fogassi, L. (2011). Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS ONE, 6*(2783), 1−10.

Cross, E. S., Hamilton, A. F. D. C., & Grafton, S. T. (2006). Building a motor simulation de novo: observation of dance by dancers. *NeuroImage, 31*(3), 1257−1267.

Darwin, C. (1872). *The Expression Of The Emotions In Man And Animals* (3rd ed.). New York: Oxford University.

Dickson, B. J. (2002). Molecular mechanisms of axon guidance. *Science,*

298(5600), 1959−1964.

Duffy, J. R. (2005). *Motor Speech Disorders* (2nd ed.). US: Elsevier.

Dum, R. P., & Strick, P. L. (1991). The origin of corticospinal projections from the premotor areas in the frontal lobe. *The Journal of Neuroscience, 11*, 667−689.

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

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164*(3875), 86−88.

Fairbanks, G., & Pronovost, W. (1938). Vocal Pitch during Simulated Emotion. *Science, 88*(2286), 382−383.

Felsenfeld, S., Kirk, K. M., Zhu, G., Statham, D. J., Neale, M. C., & Martin, N. G. (2000). A study of the genetic and environmental etiology of stuttering in a selected twin sample. *Behavior Genetics, 30*(5), 359−66.

Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press.

Fling, B. W., Benson, B. L., & Seidler, R. D. (2013). Transcallosal sensorimotor fiber tract structure-function relationships. *Human Brain Mapping*, 34(2), 384−395.

Foerster, O. (1931). The cerebral cortex in man. *Lancet, 218*(5632), 309−312.

Freeman, F., & Ushijima, T. (1978). Laryngeal muscle activity during stuttering. *Journal of Speech and Hearing Research*, 21(3), 538−562.

Gardner, B. T., & Gardner, R. A. (1985). Signs of intelligence in cross-fostered chimpanzees. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 308*, 159−176.

Geschwind, N. (1970). The organization of language and the brain. *Science, 170*(3961), 940−944.

Giret, N., Kornfeld, J., Ganguli, S., & Hahnloser, R. H. R. (2014). Evidence

for a causal inverse model in an avian cortico-basal ganglia circuit. *Proceedings of the National Academy of Sciences of the United States of America, 111*(16), 6063−6068.

Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage, 52*(3), 862−874.

Grabski, K., Lamalle, L., Vilain, C., Schwartz, J.-L., Vallée, N., Tropres, I. et al. (2012). Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Human Brain Mapping, 33*(10), 2306−2321.

Greenlee, J. D. W., Oya, H., Kawasaki, H., Volkov, I. O., Kaufman, O. P., Kovach, C. et al. (2004). A functional connection between inferior frontal gyrus and orofacial motor cortex in human. *Journal of Neurophysiology, 92*(2), 1153−1164.

Han, T.-U., Park, J., Domingues, C. F., Moretti-Ferreira, D., Paris, E., Sainz, E. et al. (2014). A study of the role of the FOXP2 and CNTNAP2 genes in persistent developmental stuttering. *Neurobiology of Disease, 59*, 23-31.

Hast, M. H., Fischer, J. M., & Wetzel, A. B. (1974). Cortical motor representation of the laryngeal muscles in macaca mulatta. *Brain, 73*, 229−240.

Hayes, C. (1952). *The ape on our house*. London: Victor Gollancz LTD.

Hayes, K. J., & Hayes, C. (1951). The Intellectual Development of a Home-Raised Chimpanzee. *Proceedings of The American Philosophical Society, 95*, 105−109.

Hernández-Morato, I., Pascual-Font, A., Ramírez, C., Matarranz-Echeverría, J., McHanwell, S., Vázquez, T. et al. (2013). Somatotopy of the neurons innervating the cricothyroid, posterior cricoarytenoid, and thyroarytenoid muscles of the rat's larynx. *Anatomical Record, 296*(3), 470−479.

Hessler, N. A., & Doupe, A. J. (1999). Social context modulates singing-related neural activity in the songbird forebrain. *Nature Neuroscience, 2*(3), 209−211.

Hofer, S., & Frahm, J. (2006). Topography of the human corpus callosum revisited–comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *NeuroImage, 32*(3), 989−994.

Hollien, H., & Moore, P. (1960). Measurements of the vocal folds during changes in pitch. *Journal of Speech, Language and Hearing Research, 3*, 157−165.

Hurst, J. A., Baraitser, M., Auger, E., Graham, F., & Norell, S. (1990). An extended family with a dominantly inherited speech disorder. *Developmental Medicine and Child Neurology, 32*, 352−355.

Ingham, R. J., Bothe, A. K., Wang, Y., Purkhiser, K., & New, A. (2012). Phonation interval modification and speech performance quality during fluency-inducing conditions by adults who stutter. *Journal of Communication Disorders, 45*(3), 198−211.

Iwatsubo, T., Kuzuhara, S., & Kanemitsu, A. (1990). Corticofugal projections to the motor nuclei of the brainstem and spinal cord in humans. *Neurology, 40*, 309−312.

Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour, 60*(1), 1−11.

Jarvis, E. (2007). Neural systems for vocal learning in birds and humans: A synopsis. *Journal of Ornithology, 148*(1), 35−44.

Jarvis, E., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., et al. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience, 6*, 151−159.

Jbabdi, S., & Johansen-berg, H. (2011). Europe PMC Funders Group Tractography - where do we go from here? *Brain Connectivity, 1*(3), 169−183.

Jürgens, U. (1974). On the elicitability of vocalization from the cortical larynx area. *Brain Research, 81*(3), 564−566.

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews, 26*(2), 235−258.

Jürgens, U. (2009). The neural control of vocalization in mammals: A review. *Journal of Voice, 23*(1), 1−10.

Jürgens, U., & Ehrenreich, L. (2007). The descending motorcortical pathway to the laryngeal motoneurons in the squirrel monkey. *Brain Research, 1148*, 90−95.

Jürgens, U., Kirzinger, A., & von Cramon, D. (1982). The effects of deep-reaching lesions in the cortical face area on phonation: A combined case report and experimental monkey study. *Cortex, 18*, 125−139.

Jürgens, U., & Ploog, D. (1970). Cerebral representation of vocalization in the squirrel monkey. *Experimental Brain Research, 554*, 532−554.

Jürgens, U., & Pratt, R. (1979a). Role of the periaqueductal grey in vocal expression of emotion. *Brain Research, 167*, 367−378.

Jürgens, U., & Pratt, R. (1979b). The cingular vocalization pathway in the squirrel monkey. *Experimental Brain Research, 510*, 499−510.

Kang, C., Riazuddin, S., Mundorff, J., Krasnewich, D., Friedman, P., Mullikin, J., & Drayna, D. (2010). Mutation in the lysosomal enzyme-targeting pathway and persistent developmental stuttering. *The New England Journal of Medicine, 362*(8), 677−685.

Kao, M. H., & Brainard, M. S. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. Journal of *Neurophysiology, 96*(3), 1441−1455.

Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005). Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature, 433*(7026), 638−643.

Kellogg, W. N. (1968). Communication and Language in the Home-Raised Chimpanzee. *Science, 162*, 423-427.

Kempster, G. B., Larson, C. R., & Kistler, M. K. (1988). Effects of electrical stimulation of cricothyroid and thyroarytenoid muscles on voice fundamental frequency. *Journal of Voice, 2*(3), 221−229.

King, S., & Sayigh, L. (2013). Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proceedings of the Royal Society B, 280*, 1−9.

Kirzinger, A., & Jürgens, U. (1982). Cortical lesion effects and vocalization in the squirrel monkey. *Brain Research, 233*, 299−315.

Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., & Lotze, M. (2010). The brain of opera singers: Experience-dependent changes in functional activation. *Cerebral Cortex, 20*(5), 1144−1152.

Kleber, B., Zeitouni, A. G., Friberg, A., & Zatorre, R. J. (2013). Experience-dependent modulation of feedback integration during singing: Role of the right anterior insula. *The Journal of Neuroscience, 33*(14), 6070−6080.

Knörnschild, M., Nagy, M., Metz, M., Mayer, F., & von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biology Letters*, 6(2), 156−159.

Koda, H., Oyakawa, C., Kato, A., & Masataka, N. (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour, 144*(6), 681−692.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech?: Vocal imitation. *Journal of the Acoustical Society of America, 100*(4), 2425−2438.

Kuypers, H. G. J. M. (1958a). Corticobulbar Connexions To the Pons and Lower Brain-Stem in Man. *Brain, 81*(3), 364−388.

Kuypers, H. G. J. M. (1958b). Some projections from the peri-central cortex to the pons and lower brain stem in monkey and chimpanzee. *The Journal of Comparative Neurology, 110*, 221−255.

Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). New York: Cambridge University Press.

Lai, C. S., Fisher, S. E., Hurst, J. a, Vargha-Khadem, F., & Monaco, a P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413*(6855), 519−23.

Lameira, A. R., Hardus, M. E., Bartlett, A. M., Shumaker, R. W., Wich, S. A., & Menken, S. B. J. (2015). Speech-Like Rhythm in a Voiced and Voiceless Orangutan Call. *PLoS ONE, 10*(1), e116136.

Lan, J., Song, M., Pan, C., Zhuang, G., Wang, Y., Ma, W. et al. (2009). Association between dopaminergic genes (SLC6A3 and DRD2) and stuttering among Han Chinese. *Journal of Human Genetics, 54*(8), 457−60.

Laukka, P., Elfenbein, H. A., Söder, N., Nordström, H., Althoff, J., Chui, W. et al. (2013). Cross-cultural decoding of positive and negative non-linguistic emotion vocalizations. *Frontiers in Psychology, 4*, 1−8.

Leyton, S., & Sherrington, C. (1917). Observations on the excitable cortex of the chimpanzee, organ-utan, and gorilla. *Experimental Physiology, 11*(2), 135−222.

Lieberman, P., & Michaels, S. B. (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *Journal of the Acoustical Society of America, 34*(7), 922−927.

Liégeois, F., Morgan, A. T., Connelly, A., & Vargha-Khadem, F. (2011). Endophenotypes of FOXP2: Dysfunction within the human articulatory network. *European Journal of Paediatric Neurology*, 15(4), 283−288.

Lisker, L., & Abramson, A. S. (1966). Some effects of context on voice onset time in English stops. *Language and Speech, 10*(1), 1−28.

Liu, W., & Nottebohm, F. (2005). Variable rate of singing and variable song duration are associated with high immediate early gene expression in two anterior forebrain song nuclei. *Proceedings of the National Academy of Sciences of the United States of America, 102*(30), 10724−10729.

Loucks, T. M. J., Poletto, C. J., Simonyan, K., Reynolds, C. L., & Ludlow, C. L. (2007). Human brain activation during phonation and exhalation: Common volitional control for two upper airway functions. *NeuroImage, 36*(1), 131−143.

Lowell, S. Y., & Story, B. H. (2006). Simulated effects of cricothyroid and thyroarytenoid muscle activation on adult-male vocal fold vibration. The *Journal*

of the Acoustical Society of America, 120(1), 386−397.

MacDougall-Shackleton, S. A, & Ball, G. F. (1999). Comparative studies of sex-differences in the song-control system of songbirds. Trends in *Neuroscience, 22*(10)(99), 432−436.

Martin, R., & Goodyear, B. (2001). Cerebral Cortical Representation of Automatic and Volitional Swallowing in Humans. *Journal of Neurophysiology, 85*, 938−950.

Martin, R., MacIntosh, B. J., Smith, R. C., Barr, A. M., Stevens, T. K., Gati, J. S., & Menon, R. S. (2004). Cerebral areas processing swallowing and tongue movement are overlapping but distinct: a functional magnetic resonance imaging study. *Journal of Neurophysiology, 92*(4), 2428−2443.

Mazzone, S. B., Cole, L. J., Ando, A., Egan, G. F., & Farrell, M. J. (2011). Investigation of the neural control of cough and cough suppression in humans using functional brain imaging. *The Journal of Neuroscience, 31*(8), 2948−2958.

Mueller, V. A., Brass, M., Waszak, F., & Prinz, W. (2007). The role of the preSMA and the rostral cingulate zone in internally selected actions. *NeuroImage, 37*(4), 1354−1361.

Myers, R. E. (1976). Comparative neurology of vocalization and speech: Proof of a dichotomy. *Annals of the New York Academy of Sciences, 280*(1), 745−757.

Netz, J., Ziemann, U., & Hömberg, V. (1995). Hemispheric asymmetry of transcallosal inhibition in man. *Experimental Brain Research, 104*(3), 527−533.

Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., & Jenner, K. C. (2000). Cultural revolution in whale songs. *Nature, 408*(6812), 537.

Nottebohm, F. (1972). The origins of vocal learning. *The American Naturalist, 106*(947), 116−140.

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, Serinus canarius. *The Journal of Comparative Neurology, 165*(4), 457−486.

Olthoff, A., Baudewig, J., Kruse, E., & Dechent, P. (2008). Cortical sensorimotor control in vocalization: A functional magnetic resonance imaging study. *The Laryngoscope, 118*(11), 2091−2096.

Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. American *Journal of Primatology, 73*(6), 530−544.

Pabst, F., & Sundberg, J. (1993). Tracking multi-channel electroglottograph measurement of larynx height in singers. *Logopedics Phonatrics Vocology, 18*, 143−152.

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representations in the cerebral cortex of man as studied by electrical stimulation. *Brain, 60*, 389−443.

Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience, 4*, 1−24.

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L. et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science, 346*(6215), 1256846−1256846.

Picard, N., & Strick, P. L. (1996). Motor areas of the medial wall: A review of their location and functional activation. *Cerebral Cortex, 6*(3), 342−353.

Pierce, J. D. (1985). A review of attempts to condition operantly alloprimate vocalizations. *Primates, 26*(2), 202−213.

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Elephants are capable of vocal learning. *Nature, 434*, 455−456.

Ralls, K., Fiorelli, P., & Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals, Phoca vitulina. *Canadian Journal of Zoology, 63*, 1050−1056.

Ramsay, S. C., Adams, L., Murphy, K., Corfield, D. R., Grootoonk, S., Bailey, D. L. et al. (1993). Regional cerebral blood flow during volitional expiration in man: A comparison with volitional inspiration. *The Journal of Physiology,*

461(1993), 85−101.

Riebel, K. (2003). The "Mute" Sex Revisited: Vocal Production and Perception Learning in Female Songbirds. *Advances in the Study of Behavior, 33*, 49−86.

Rödel, R. M. W., Olthoff, A., Tergau, F., Simonyan, K., Kraemer, D., Markus, H., & Kruse, E. (2004). Human cortical motor representation of the larynx as assessed by transcranial magnetic stimulation (TMS). *The Laryngoscope, 114*(5), 918−22.

Roper, A., & Zann, R. (2006). The Onset of Song Learning and Song Tutor Selection in Fledgling Zebra Finches. *Ethology, 112*(5), 458−470.

Roubeau, B., Chevrie-Muller, C., & Saint Guily, J. (1997). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Otolaryngologica, 117*(3), 459−464.

Sanvito, S., Galimberti, F., & Miller, E. H. (2007). Observational evidences of vocal learning in southern elephant seals: A longitudinal study. *Ethology, 113*(2), 137−146.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology, 63*(11), 2251−2272.

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2004). The influence of hearing impairment on preverbal emotional vocalizations of infants. *Folia Phoniatrica et Logopaedica, 56*(1), 27−40.

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2006). Vocal expression of emotions in normally hearing and hearing-impaired infants. *Journal of Voice*, 20(4), 585−604.

Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology, 66*(2), 310−328.

Schmidt, M. F., Ashmore, R. C., & Vu, E. T. (2004). Bilateral control and interhemispheric coordination in the avian song motor system. *Annals of the*

New York Academy of Sciences, 1016, 171−186.

Schulz, G. M., Varga, M., Jeffires, K., Ludlow, C. L., & Braun, a R. (2005). Functional neuroanatomy of human vocalization: An H215O PET study. *Cerebral Cortex, 15*(12), 1835−1847.

Seikel, J., King, D., & Drumwright, D. (2010). Anatomy of phonation. In *Anatomy & Physiology for Speech, Language, and Hearing* (4th ed., pp. 165−222). US: Delmar, Centage Learning.

Shima, K. (1998). Role for Cingulate Motor Area Cells in Voluntary Movement Selection Based on Reward. *Science, 282*(5392), 1335−1338.

Shipp, T., & Izdebski, K. (1975). Vocal frequency and vertical larynx positioning by singers. *Journal of the Acoustical Society of America, 58*(5), 1104−1106.

Simon-Thomas, E. R., Keltner, D. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion, 9*(6), 838−846.

Simonyan, K., & Jürgens, U. (2002). Cortico-cortical projections of the motorcortical larynx area in the rhesus monkey. *Brain Research, 949*(1-2), 23−31.

Simonyan, K., & Jürgens, U. (2003). Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Research, 974*(1-2), 43−59.

Simonyan, K., & Jürgens, U. (2005). Afferent subcortical connections into the motor cortical larynx area in the rhesus monkey. *Neuroscience, 130*(1), 119−131.

Simonyan, K., Ostuni, J., Ludlow, C. L., & Horwitz, B. (2009). Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *The Journal of Neuroscience, 29*(47), 14912−14923.

Simonyan, K., Saad, Z. S., Loucks, T. M. J., Poletto, C. J., & Ludlow, C. L. (2007). Functional neuroanatomy of human voluntary cough and sniff production. *NeuroImage, 37*(2), 401−409.

Sohrabji, F., Nordeen, E. J., & Nordeen, K. W. (1990). Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. *Behavioral and Neural Biology, 53*(1), 51−63.

Stoeger, A. S., Mietchen, D., Oh, S., de Silva, S., Herbst, C. T., Kwon, S., & Fitch, W. T. (2012). An Asian elephant imitates human speech. *Current Biology, 22*(22), 2144−2148.

Sutton, D., Larson, C., & Lindeman, R. C. (1974). Neocortical and limbic lesion effects on primate phonation. *Brain Research, 71*(1), 61−75.

Sutton, D., Trachy, R. E., & Lindeman, R. C. (1981). Primate phonation: Unilateral and bilateral cingulate lesion effects. *Behavioural Brain Research, 3*(1), 99−114.

Takai, O., Brown, S., & Liotti, M. (2010). Representation of the speech effectors in the human motor cortex: Somatotopy or overlap? *Brain and Language, 113*(1), 39−44.

Titze, I. R., Luschei, E. S., & Hirano, M. (1989). Role of the thyroarytenoid muscle in regulation of fundamental frequency. *Journal of Voice, 3*(3), 213−224.

Tumer, E. C., & Brainard, M. S. (2007). Performance variability enables adaptive plasticity of "crystallized" adult birdsong. *Nature, 450*(7173), 1240−1244.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage, 16*(3), 765−780.

Vernes, S. C., Oliver, P. L., Spiteri, E., Lockstone, H. E., Puliyadi, R., Taylor, J. M. et al. (2011). Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genetics, 7*(7), e1002145.

Vernes, S. C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J. M., Davies, K. E. et al. (2007). High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *American Journal of Human Genetics, 81*(6), 1232−1250.

Vilkman, E., Sonninen, A., Hurme, P., & Körkkö, P. (1996). External laryngeal frame function in voice production revisited: A review. *Journal of Voice, 10*(1), 78−92.

Vogt, B. A., Berger, G. R., & Derbyshire, S. W. G. (2003). Structural and functional dichotomy of human midcingulate cortex. *European Journal of Neuroscience, 18*, 3134−3144.

Vogt, B. A., Vogt, L., Farber, N. B., & Bush, G. (2005). Architecture and neurocytology of monkey cingulate gyrus. *Journal of Computational Neuroscience, 485*(3), 218−239.

Wahl, M., Lauterbach-Soon, B., Hattingen, E., Jung, P., Singer, O., Volz, S. et al. (2007). Human motor corpus callosum: Topography, somatotopy, and link between microstructure and function. *The Journal of Neuroscience, 27*(45), 12132−12138.

Watkins, K. E., Dronkers, N. F., & Vargha-Khadem, F. (2002). Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain, 125*, 452−464.

Watkins, K. E., Smith, S. M., Davis, S., & Howell, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. *Brain, 131*, 50−59.

Watkins, K. E., Vargha-Khadem, F., Ashburner, J., Passingham, R. E., Connelly, A., Friston, K. J. et al. (2002). MRI analysis of an inherited speech and language disorder: Structural brain abnormalities. *Brain, 125*, 465−478.

Watkins, K., & Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience, 16*(6), 978−987.

Wattendorf, E., Westermann, B., Fiedler, K., Kaza, E., Lotze, M., & Celio, M. R. (2013). Exploration of the neural correlates of ticklish laughter by functional magnetic resonance imaging. *Cerebral Cortex, 23*, 1280−1289.

Wich, S. a, Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates, 50*(1), 56−64.

Wild, J. M., Williams, M. N., & Suthers, R. A. (2000). Neural pathways for bilateral vocal control in songbirds. *The Journal of Comparative Neurology, 426*, 413−426.

Wu, S.-X., Goebbels, S., Nakamura, K., Nakamura, K., Kometani, K., Minato, N. et al. (2005). Pyramidal neurons of upper cortical layers generated by NEX-positive progenitor cells in the subventricular zone. *Proceedings of the National Academy of Sciences of the United States of America, 102*(47), 17172−17177.

Wymbs, N. F., Ingham, R. J., Ingham, J. C., Paolini, K. E., & Grafton, S. T. (2013). Individual differences in neural regions functionally related to real and imagined stuttering. *Brain and Language, 124*(2), 153−164.

Yip, M. (2002). Tone. New York, NY: Cambridge.

Yoshida, Y., Tanaka, Y., Saito, T., Shimazaki, T., & Hirano, M. (1992). Peripheral nervous system in the larynx. *Folia Phoniatrical, 44*, 194−219.

Appendix A

Supplement to Chapter 2

Table A.1: First Supplement to Chapter 2.

Description of studies included in the meta-anlalysis of affective and linguistic prosody perception.

Study	Affective Prosody	Linguistic Prosody	N	Experimental Task	Control Task	Verbal Working Memory Demands	Functional Coverage	Stereotaxic Space	Stimulus Language	Imaging Modality
Gandour et al., 2003b (English speakers)	Y	Y	10	Same/Different	Passive listening	High	Whole Brain	Tal	Mandarin	fMRI
Gandour et al., 2003b (Mandarin speakers)	Y	Y	10	Same/Different	Passive listening	High	Whole Brain	Tal	Mandarin	fMRI
Bach et al., 2008	Y	N	16	Discrimination	Neutral, Gender	Low	Partial	MNI	German	fMRI
Beaucousin et al., 2007	Y	N	23	Discrimination	Grammar, Synth	Low	Partial	MNI	French	fMRI
Bruck et al., 2011	Y	N	24	Discrimination	Semantics, Vowels	Low	Whole Brain	MNI	German	fMRI
Buchanan et al., 2000	Y	N	10	Detect target	Phonetics	Low	Whole Brain	Tal	English	fMRI
Ethofer et al., 2006	Y	N	24	Discrimination	Rating, Semantics	Low	Partial	MNI	German	fMRI
Ethofer et al., 2009	Y	N	24	Discrimination	Neutral, Word Class	Low	Whole Brain	MNI	German	fMRI
Fruzholz et al., 2011	Y	N	17	Discrimination	Neutral	Low	Whole Brain	MNI	French	fMRI
George et al., 1996	Y	N	13	Discrimination	Semantics, Repetition	Low	Whole Brain	Tal	English	PET
Imaizumi et al., 1997	Y	N	6	Discrimination	Speaker Identity	Low	Whole Brain	Tal	Japanese	PET
Johnstone et al., 2006	Y	N	34	Discrimination	Other emotions	Low	Whole Brain	MNI	English	fMRI
Kotz et al., 2003	Y	N	12	Discrimination	Neutral	Low	Partial	Tal	German	fMRI
Leitman et al., 2010	Y	N	20	Discrimination	Neutral	Low	Partial	MNI	English	fMRI
Mitchell et al., 2003	Y	N	13	Attend to prosody	Attend to semantics	Low	Partial	Tal	English	fMRI
Quadflieg et al., 2008	Y	N	12	Discrimination	Neutral	Low	Partial	Tal	German	fMRI
Szameitat et al., 2010	Y	N	18	Discrimination	Count bouts	Low	Partial	MNI	German	fMRI
Wildgruber et al., 2002	Y	N	12	Identification, Relative expressiveness	2nd>1st stimulus	High	Partial	MNI	German	fMRI
Wildgruber et al., 2005	Y	N	10	Discrimination	Vowel discrimination	Low	Whole Brain	MNI	German	fMRI
Doherty et al., 2004	N	Y	11	Discrimination	Statements, Questions	Low	Partial	Tal	English	fMRI
Gandour et al., 2003a (English speakers)	N	Y	10	Same/Different	Passive listening	High	Whole Brain	Tal	Mandarin	fMRI
Gandour et al., 2003a (Mandarin speakers)	N	Y	10	Same/Different	Passive listening	High	Whole Brain	Tal	Mandarin	fMRI
Geiser et al., 2008	N	Y	24	Speech rhythm discrimination	Intonation Discrimination	Low	Partial	MNI	German	fMRI
Klein et al., 2011	N	Y	24	Same/Different	Vowel discrimination	High	Partial	MNI	German	fMRI
Meyer et al., 2004 N	Y	14	Same/Different	Monotonous speech	High	Partial	Tal	German	fMRI	
Meyer et al., 2002	N	Y	14	Discrimination	Synctatic speech	Low	Partial	Tal	German	fMRI
Strelnikov et al., 2006	N	Y	12	Discrimination	Non-Segmented speech	Low	Partial	MNI	Russian	PET

Table A.2: Second supplement to Chapter 2.

Complete bibliography for the meta-analysis of affective and linguistic prosody.

Bach, D.R., Grandjean, D., Sander, D., Herdener, M., Strik, W.K., and Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. Neuroimage, 42(2), 919-27.

Beaucousin, V., Lacheret, A., Turbelin, M.-R., Morel, M., Mazoyer, B., and Tzourio-Mazoyer, N. (2007). fMRI study of emotional speech comprehension. Cerebral Cortex, 17(2), 339-52.

Brck, C., Kreifelts, B., Kaza, E., Lotze, M., and Wildgruber, D. (2011). Impact of personality on the cerebral processing of emotional prosody. Neuroimage, 58(1), 259-68.

Buchanan, T.W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K. et al., (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. Cognitive Brain Research, 9(3), 227-38.

Doherty, C.P., West, W.C., Dilley, L.C., Shattuck-Hufnagel, S., and Caplan, D. (2004). Question/statement judgments: An fMRI study of intonation processing. Human Brain Mapping, 23(2), 85-98.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J. et al., (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. Neuroimage, 30(2), 580-7.

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., and Vuilleumier, P. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. Journal of Cognitive Neuroscience, 21(7), 1255-68.

Frhholz, S., Ceravolo, L., and Grandjean, D. (2011). Specific brain networks during explicit and implicit decoding of emotional prosody. Cerebral Cortex, 22(5), 1107-17.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L. et al.,

(2003a). Temporal integration of speech prosody is shaped by language experience: An fMRI study. Brain and Language, 84(3), 318-36.

Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., and Li, X. (2003b). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. Human Brain Mapping, 18(3), 149-57.

Geiser, E., Zaehle, T., Jancke, L., and Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by metrical speech processing. Journal of Cognitive Neuroscience, 20(3), 541-52.

George, M., Parekh, P., and Rosinsky, N. (1996). Understanding emotional prosody activates right hemisphere regions. Archives of Neurology, 53, 665-70. Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., and Fukuda, H. et al., (1997). Vocal identification of speaker and emotion activates different brain regions. Neuroreport, 8(12), 2809-12.

Johnstone, T., van Reekum, C.M., Oakes, T.R., and Davidson, R.J. (2006). The voice of emotion: An fMRI study of neural responses to angry and happy vocal expressions. Social Cognitive and Affective Neureuscience, 1(3), 242-9.

Klein, E., Domahs, U., Grande, M., and Domahs, F. (2011). Neuro-cognitive foundations of word stress processing: Evidence from fMRI. Behavioral and Brain Functions, 7, 1-17.

Kotz, S. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. Brain and Language, 86(3), 366-76. Leitman, D.I., Wolf, Daniel, H., Ragland, J.D., Laukka, P., Loughead, J. et al., (2010). Its not what you say, but how you say it: A reciprocal temporo-frontal network for affective prosody. Frontiers in Human Neuroscience, 4, 1-13.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G., and von Cramon, D.Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Human Brain Mapping, 17(2), 73-88.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A.D., and von Cramon, D.Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. Brain and Language, 89(2), 277-89.

Mitchell, R.L.C., Elliott, R., Barry, M., Cruttenden, A., and Woodruff, P.W.R.

(2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. Neuropsychologia, 41(10), 1410-21.

Quadflieg, S., Mohr, A., Mentzel, H.-J., Miltner, W.H.R., and Straube, T. (2008). Modulation of the neural network involved in the processing of anger prosody: The role of task-relevance and social phobia. Biological Psychology, 78(2), 129-37.

Strelnikov, K.N., Vorobyev, V.A, Chernigovskaya, T.V., and Medvedev, S.V. (2006). Prosodic clues to syntactic processing-a PET and ERP study. Neuroimage, 29(4), 1127-34.

Szameitat, D.P., Kreifelts, B., Alter, K., Szameitat, A.J., Sterr, A., Grodd, W. et al., (2010). It is not always tickling: Distinct cerebral responses during perception of different laughter types. Neuroimage, 53(4), 1264-71.

Wildgruber, D., Pihan, H., Ackermann, H., Erb, M., and Grodd, W. (2002). Dynamic brain activation during processing of emotional intonation: Influence of acoustic parameters, emotional valence, and sex. Neuroimage, 15(4), 856-69.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T. et al., (2005). Identification of emotional intonation evaluated by fMRI. Neuroimage, 24(4), 1233-41.

Appendix B

Supplement to Chapter 3

Table B.1: First Supplement to Chapter 3.

Discrimination accuracy in the stimulus validation experiment. The mean and standard deviation of the accuracy scores are listed for recordings of each emotion type. The stimuli with the highest discrimination scores were taken as the best available exemplars of vocalizations expressing each emotion, and were therefore reserved for the neuroimaging experiment.

Stimulus set	Happiness	Pleasure	Sadness	Disgust	Overall
Training	0.70 (SD 0.10)	0.57 (SD 0.34)	0.63 (SD 0.22)	0.72 (SD 0.98)	0.66 (SD 0.20)
Experiment	0.76 (SD 0.09)	0.83 (SD 0.09)	0.73 (SD 0.13)	0.77 (SD 0.08)	0.76 (SD 0.11)

Table B.2: Second supplement to Chapter 3.

Region-of-interest analysis for the periaquaductal gray are invariant to the size of the ROI.

5mm ROI	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Task	1	1.575	1.5748	6.502	0.0148*
Content	1	0.047	0.0473	0.195	0.661
Task:Content	1	0.112	0.1119	0.462	0.5007
Residuals	39	9.446	0.2422		

3mm ROI	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Task	1	1.559	1.5587	6.372	0.0158*
Content	1	0.093	0.093	0.38	0.5411
Task:Content	1	0.22	0.2199	0.899	0.3489
Residuals	39	9.54	0.2446		

1mm ROI	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Task	1	1.336	1.3357	5.846	0.0204*
Content	1	0.257	0.2573	1.126	0.2952
Task:Content	1	0.37	0.3695	1.617	0.211
Residuals	39	8.911	0.2285		

Appendix C

Supplement to Chapter 5

Table C.1: First Supplement to Chapter 5.

This table lists descriptive information about data and data handling for studies that contributed to the meta-analysis of stuttering. Asterisks mark fMRI studies that utilized sparse or clustered event-related designs to eliminate scanner noise during speech tasks.

	Sample Size			Original			
Study	Stuttering	Control	Modality	Space	Software	Language	Participant Screening
Braun et al. (1997)	18	20	PET	TAL	SPM	English	Neuropsychiatric illness
Neumann et al. (2003)	16	16	fMRI	MNI	SPM	German	
Priebishch et al. (2003)	16	16	fMRI	MNI	SPM	German	
Giraud et al. (2008)	16	16	fMRI	MNI	SPM	German	
De Nil et al. (2008)	15	15	fMRI*	MNI	SPM	English	
Chang et al. (2009)	20	20	fMRI*	MNI	AFNI	English	Neurological disorders
Kell et al. (2009)	13	13	fMRI	MNI	SPM	German	
Sakai et al. (2009)	8	10	fMRI*	TAL	SPM	Japanese	
Lu et al. (2010)	10	9	fMRI	TAL	AFNI	Chinese	
Howell et al. (2012)	9	9	fMRI	MNI	AFNI	Chinese	Neurological disorders
Fox et al. (2000)	10	10	fMRI	TAL	MIPS	English	
Ingham et al. (2004)	10	10	PET	TAL	SPM	English	
Toyomura et al (2011)	12	12	fMRI*	MNI	SPM	Japanese	"Speech, language & hearing"
Ingham et al. (2012)	18	12	PET	MNI	SPM	English	Neurological disorder
Jiang et al. (2012)	20	0	fMRI*	MNI	AFNI	Chinese	Psychiatric/Neurological disorders
Wymbs et al (2013)	1/1/1/1	0	fMRI	TAL	SPM	English	
den Ouden et al. (2013)	1	0	fMRI*	MNI	SPM	English	