UNIVERSITAT DE BARCELONA

The Role of Vocal Learning in Language. Evolution and Development

Qing Zhang

The Role of Vocal Learning in Language Evolution and Development

Qing Zhang

A Doctor Dissertation Submitted
in Partial Fulfilment of the
Requirements of the Degree of
Doctor of Philosophy
to the Doctor Program
Cognitive Science and Language
Department of Catalan Philology and General linguistics
Universitat de Barcelona

under the supervision of

Joana Roselló Ximenes

Universitat de Barcelona

UNIVERSITAT DE BARCELONA

June 2017

Table of content

Abstract

Vocal learning, one of the subcomponents of language, is put at center stage in this dissertation. The overall hypothesis is that vocal learning lays the foundation for both language evolution (phylogeny) and development (ontogeny), and also high-level cognition. The computational ability found in vocal learning is seen as so enhanced in humans as to yield the kind of recursion that supports language. Empirical evidence on vocal learning in nonhuman animals and humans from behavioral, neuroanatomical, neurophysiological, genetic, and evolutionary fields is suggestive that vocal learning interacts with other cognitive domains at multiple levels. The positive correlation between the hippocampal volume and open-ended vocal production in avian vocal learning species suggests the possible involvement of the hippocampus in vocal learning. The empirical studies of foxp2 in nonhuman animals and humans suggest that foxp2 plays a role in multimodal communication and general cognition.

Phylogenetically, Sapiens' vocal learning abilities are unique among primates. Compared with nonhuman primates, our species possesses stronger and more enhanced connections between the superior temporal cortex and premotor cortex as well as the striatum. In Sapiens, meaning aside, vocal learning as such can explain many features found in speech and its ontogeny such as the specialized auditory mechanism for speech, the preferential attention to speech in newborns, the primacy of vocal imitation among multimodal (visual and auditory) imitative skills and the stages seen in learning to speak.

All these characteristics seem to be different and abnormal, albeit to different degrees, in autism. A 25-30% of the autistic population is non/minimally verbal but even the high functioning end of the autistic spectrum presents with abnormalities, such as difficulties in processing speed and an impaired imitative capacity that could be satisfactorily explained if language entered again the definition (and diagnosis) of what autism is, with an special emphasis on vocal learning.

Resum

El *vocal learning*, un dels subcomponents del llenguatge, ocupa un espai central en aquesta tesi. La hipòtesi general és que el *vocal learning* constitueix el fonament de l'evolució (filogènia) i del desenvolupament (ontogènia) lingüístics, i també de la cognició. L'habilitat computacional que es dóna en el *vocal learning* es veu en els humans tan potenciada com per ser la base del tipus de recursió en què es basa el llenguatge. Proves empíriques sobre el *vocal learning* en animals no humans i en humans, des de camps que inclouen des del comportament, la neuroanatomia, la neurofisiologia, la genètica i la teoria de l'evolució, suggereixen que el *vocal learning* interactua amb altres dominis cognitius a molts i diferents nivells. La correlació positiva entre el volum de l'hipocamp i el caràcter *open-ended* de la producció vocal en espècies d'aus amb *vocal learning* apunta a una possible contribució de l'hipocamp en el *vocal learning*. Els estudis empírics sobre el foxp2 en animals no humans i en humans suggereix que el foxp2 juga un paper en la comunicació multimodal i la cognició.

Filogenèticamet, les habilitats de *vocal learning* en el Sapiens són úniques entre els primats. Comparada amb els primats no humans, la nostra espècie posseeix unes connexions més denses i potents entre el còrtex temporal superior i el còrtex premotor així com l'estriat. En el Sapiens, deixant de banda el significat, el *vocal learning* tot sol pot explicar molts trets de la parla i la seva ontogènia com ara l'especialització auditiva per a la parla, l'atenció preferent a la parla en els nadons, la primacia de la imitació vocal entre les habilitats imitatives multimodals (de base visual i auditiva), i els estadis que s'observen en l'adquisició de la parla.

Totes aquestes característiques sembla que són diferents i anòmales, tot i que en

diferent graus, en l'autisme. Un 25-30% de la població autista és no verbal o mínimament però fins i tot a la banda de l'espectre autista que es considera d'alt funcionament s'hi donen anomalies, tal com ara un cert dèficit en velocitat de processament i una capacitat deficient d'imitació, que podrien explicar-se més satisfactòriament si un dèficit de llenguatge entrés altra vegada a la definició (i diagnòstic) del que és l'autisme, amb un èmfasi especial en el *vocal learning*.

Initial fall in the journey

Crystallized landscape and sky

All refreshing and energetic

Nerves persuaded me to try

The tenderness melted the spirit

Hiding the darkness under the sunlight

Emotion manipulated the eyes

Life always refuses regrets

Persistent on the way witnessing the passage of time

Sep. 18, 2016

Acknowledgments

It has been a long journey with endeavor and perspiration, but considerable pleasure and satisfaction. This part is dedicated to the people whose presence has been indispensable for me to finish my dissertation.

I would like to express my profound gratitude to my supervisor, Joana Rosselló. She has been the sunlight appearing in the darkness. Although we have only worked together for one and a half years, what we have discussed and what she has taught me have had a big impact on me not only in my research, but also in my life. I cannot imagine how and where I can arrive so far without her.

My thanks also go to my former supervisor, Cedric Boeckx. Working with him has made me realize the importance of being independent in academia. I have learned a lot at intellectual level for the two years when he always reminded me to be curious and suspicious to anything.

Evelina Leivada and Pedro Martins deserve special thanks. They have helped me a lot both in my research and in my life since I moved to Barcelona.

I thank my friend and colleague Ruoyang Shi, who is always supporting me since I entered the door of linguistics for the past seven years. With his presence in Barcelona, the office is always full of laughs. I also thank my friend and colleague Yuliang Sun, who keeps giving me advice in my life and cheering me up when I was upset.

I am grateful to my best friend Xiao Han. We have known each other for fourteen years. Her presence makes it so much easier for me to live in Barcelona. Her

continuous encouragement and consolation give me reasons to be stronger. Congratulate to her having a baby recently! I am also indebted to my friend Xiaomeng Wang, whose company has been the biggest spiritual reliance for me at the beginning of my living in Barcelona.

I am heartedly appreciative to my boyfriend, Jian Lin. His appearance enlightens my life. The moments, the laughs, the tears, the dilemmas, make me feel complete. His ongoing invigoration provides me with abundance of courage and confidence in my study and life. I deeply feel like family being with him and his family. It has been being the most invaluable experience in my life.

No word can describe my gratitude to my parents. They have sacrificed more than a lot to raise me up. Their blindness does not prevent them from being fearless to let me go for my dream overseas. Their bravery and strength is the source for me to be insistent and persistent. If I could, I would be your daughter again for the next life. I love you both!

List of abbreviations

AAC	central nucleus of the anterior arcropallium
ACM	caudal medial acropallium
ASC	autism spectrum condition
AF	the arcuate fasciculus
Ai	intermediate arcopallium
AIM	Active Imitation Matching
Area X	areaX of the striatum
ASD	autism spectrum disorder
ASL	the associative sequence learning
BA	brodmann area
BVI	the Basic Vocal Imitation
CAS	childhood apraxia of speech
CDFE	cortical dysplasia-focal epilepsy
CI	conceptual-intentional system
CM	caudal mesopallium
CMM	caudal medial mesopallium
DLH	d,l-homocysteic acid
DLM	medial nucleus of dorsolateral thalamus
DMM	magnocellular nucleus of the dorsomedial thalamus
DM	dorsal medial nucleus of the midbrain
DSM	Diagnostic and Statistical Manual
DTI	diffusion-MRI tractography
DVD	developmental verbal dyspraxia

FLB	the faculty of language in broad sense
FLN	the faculty of language in narrow sense
GPi	the internal globus pallidus
HCF	Hauser, Chomsky and Fitch (2002)
HVC	a letter-based name
H.M.	Henry Molaison
IEG	immediate early gene
IFG	inferior frontal gyrus
IPL	inferior parietal lobule
IT	Inspection Time
iTG	the inferior temporal gyrus
KFn	kolliker-fuse nuclei
LMAN	lateral part of MAN
LMC	laryngeal motor cortex
L1	field L1
L2	field L2
L3	field L3
MAN	magnocellular nucleus of anterior nidopallium
mPFC	the medial prefrontal cortex
mTG	the middle temporal gyrus
MMst	magnocellular nucleus of the anterior striatum
MRI	Magnetic resonance imaging
NAO	oval nucleus of the anterior nidopallium
NCM	caudal medial nidopallium

NLC	central nucleus of anterior nidopallium
nIIts	tracheosyringeal subdivision of the hypoglossal nucleus
PAG	periaqueductal grey
PDD-NOS	pervasive developmental disorder not otherwise specified
PECS	Picture Exchange Communication System
pITL	the posterior inferior temporal lobe
PMv	the ventral premotor cortex
PSI	Processing Speed Index
RA	robust nucleus of the acropallium
SM	sensory-motor system
SLF	the superior longitudinal fasciculus
SLI	specific language impairment
Spt	the posterior Sylvian fissure at the parietal-temporal boundary
STG	superior temporal gyrus
STS	the superior temporal sulcus
SVZ	subventricular zone
SWS	slow wave sleep
TD	typical development
UG	universal grammar
VA	vocal nucleus of the acropallium
VLN	vocal nucleus of the lateral nidopallium
VI	vocal imitation
WISC	Wechsler Intelligence Scale for Children

List of figures

List of tables

For my father

When I am down and oh my soul, so weary;
When troubles come and my heart burdened be;
Then I am still and wait here in the silence,
Until you come and sit awhile with me
You raise me up, so I can stand on mountains;
You raise me up, to walk on stormy seas;
I am strong, when I am on your shoulders;
You raise me up, to more than I can be
----Brendan Graham

1 Introduction

Chomsky's revolutionary assumption that language is a biological product of the human brain, with a large proportion of it being innate, so as to account for the fact that the acquisition of any language is effortless and for the infiniteness of linguistic expressions, has been the doctrine of Generative Grammar. In a series of writings by Chomsky since HCF (i.e. Hauser et al. 2002) (e.g. Hauser et al., 2014; Berwick & Chomsky, 2016; Everaert et al., 2017), how language could have evolved in human lineage has been increasingly discussed in this framework. Thus, for the first time evolutionary considerations have had an impact in shaping the so-called minimalist program, the most recent incarnation of a formal model of the architecture of language in this framework.

In the Chomskyan view, recursion has been regarded as the hallmark of the computational system of human language and is the only key element properly evolved in humans for language. In HCF terms, the Faculty of Language in the Narrow sense (FLN), which consists essentially of recursion, would be the only part of the Faculty of Language in the Broad sense (FLB) that is unique to humans. HCF however concede that this uniqueness or FLN character of recursion obtains only if the domain under consideration is that of animal communication systems. They concede at the end of the paper that other animals could possess recursion for the purposes of spatial navigation. This demotion of FLN as "unique to humans" in the confinement of animal communication is however not entirely coherent with

Chomsky’s own view that has always related recursion to thought or the conceptual-intentional (CI) system and not to the sensorimotor (SM) system which is the one required for communication. Recursion + CI qualify as primary and the SM system is considered ancillary or secondary (figure 1 A). Equivalently, language evolved for thought and not for communication.

In this dissertation we will take a different angle which does not present the inconsistencies pointed out above and will argue that language has evolved bottom-up, that is from a SM system instead of deriving from a recursive CI (figure 1 B). The SM system for language is in turn based on an ancestral vocal learning capacity. The computational mechanism for recursion in language is conceived of as an enhancement of the computational abilities associated with vocal learning as found in other nonhuman animals. The current work will then focus on the role of vocal learning in language and domain general cognition. To be more specific, we hypothesize that vocal learning serves as a foundational basis for both language development (ontogeny) and language evolution (phylogeny), and further high-level cognition.

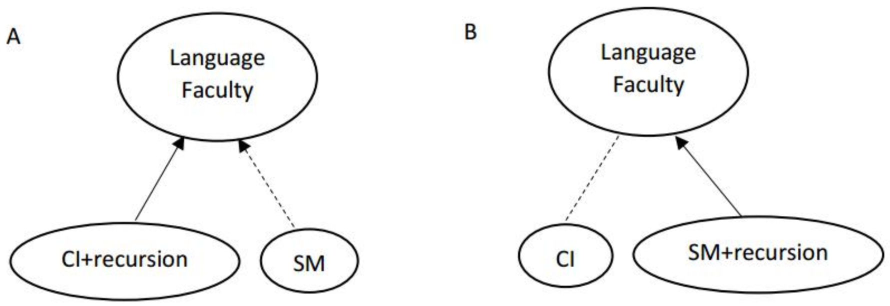


Figure 1: A. Language evolved from the conceptual-intentional (CI) system (Chomskyan view); B. Language evolved from a sensorimotor (SM) system (our view).

In fields other than linguistics, vocal learning has been intensely studied in other nonhuman species. Since it is a rare trait in animal kingdom, only present in three groups of birds (songbirds, parrots and hummingbirds), some species of mammals (bats, elephants, cetaceans, pinnipeds) including humans, researchers naturally are curious about the nature of this trait and how this trait emerged in evolution. Studies on behaviors, neuroanatomy, neurophysiology, genetics and so on, on vocal learning birds have proliferated. Prominent scholars in this field, like Jarvis, Scharff, White, Wada, Okanoya, etc., have made important contributions to vocal learning in general (e.g. Jarvis, 2007; 2009; Pektov & Jarvis, 2012; Chakraborty & Jarvis, 2015; Pfenning et al., 2014; Wang et al., 2015; Scharff & Petri, 2011; Haesler et al., 2007; Condro & White, 2014; Chen et al., 2013; Liu et al., 2013; Feenders et al., 2008; Matsunaga & Okanoya, 2014) during the last two decades. Furthermore, since vocal learning has been accepted as one of the necessary abilities for humans to acquire speech/language, comparative studies shedding light on human speech have also been productive. However, few studies have directly addressed the role of vocal learning in both language development (ontogeny) and language evolution (phylogeny), and even high-level cognition. The current dissertation presents the hypothesis that vocal learning plays an essential role in both language development (ontogeny) and language evolution (phylogeny), and high-level cognition. We will present supportive evidence from the aspects of behavior, neuroanatomy, neurophysiology, genetics, and evolution in both nonhuman animals and human.

The whole dissertation is organized as follows. In chapter two, we will present biological basis of vocal learning in both nonhuman animals and human beings from various aspects and across multiple levels. We introduce the developmental stages of vocal learning including auditory, sensorimotor and motor phases; the role of declarative and procedural memory in vocal learning; the neural pathways containing an anterior pathway for learning and a posterior pathway for production of vocal learning; the genetic basis of vocal learning; and the proposed evolutionary trajectory of vocal learning. In the meanwhile, we put forward several hypotheses in this chapter:

- There is a subsequent evolutionary order of vocal learning pathways, namely the posterior pathway evolved before and served as a prerequisite of the anterior pathway.
- The hippocampus could play an important role in the memorization phase of vocal learning and in the extended song learning in open-ended avian species, which could be analogous to second language learning.
- The two mutations of human version of FOXP2 play separate roles in the evolution of vocal learning pathways, being one for the morphological

changes for the anterior pathway and the other potentially promoted the enlargement of brain volume.

- Foxp2 is better understood as a gene that plays an essential role in a general mechanism that is underlying human cognition across domains.
- Breathing could have played a crucial role in the evolution of multimodal communication.

In chapter three, we focus on the role of vocal learning in human language and general cognition. The sounds of speech constitute the conspecifics' sounds in humans. The sounds of speech are learned in the same way as conspecifics' sounds are learned by vocal learners. A difference, however stands out, which is that there is meaning bound to the tones. This hugely consequential difference, however, does not preclude that speech processing (speech production and comprehension) and acquisition is to a great extent the same as found in nonhuman versions of vocal learning. Thus, our main proposal in chapter three is that vocal learning serves as the foundational basis for language development and evolution, as well as high-level cognition. We will address the issue of recursion and propose that the core computational ability of recursion of language could have derived from preexisting computational abilities present by definition in vocal learning. By reviewing relevant empirical studies on speech/language processing, we argue that speech is special in both perception and production developmentally as corresponds to an evolutionary specialization for sound processing in vocal learners in general. By comprehensively comparing two main theories on imitation, namely the associative sequential learning

(ASL) and the Active Intermodal Mechanism (AIM), we contend that vocal imitation is the basic one so that a Basic Vocal (BVI) deserves to be explored as the potential basis for the general imitative abilities seen in humans. The multimodality of language evolution will also be presented in this chapter. We also come up with several hypotheses:

- POU3F2 could have been one of the key factors for the emergence of recursion in human language.
- The stronger and more enhanced connectivity of the temporal lobe with other brain areas could have been the reason why auditory-vocal modality is predominant in the evolution of human language.
- The subdivision of human ventral premotor cortex (BA6) could have been the driving force for the enlargement of primary motor cortex, which gives rise to the direct corticolaryngeal connection in human lineage.
- Language evolved multimodally with auditory-vocal modality having primacy..
- Dancing and beat gesture may be traced back to the same origin.
- Vocal imitation ability is the core of multimodal and multidimensional imitative skills found in humans.

By focusing on one of the atypically cognitive developing populations, namely that of the individuals with autism spectrum disorders (ASD), chapter four revives the language account for ASD phenotypes. We propose that the problems that ASD individuals present with is originated from deficits in their vocal learning abilities as suggested by the following considerations:

- Processing sequential, fast and transient stimuli is more difficult or impaired in ASD.
- Empirical evidence from behavior, neuroanatomy and neurophysiology has shown that ASD children are deficient in both speech perception and production with supportive evidence from genetics (CNTNAP2) and neural correlates.
- The sensorimotor speech deficits could be derived from the atypical development of dorsal pathway, which is responsible for auditory-vocal integration.
- Repetitive behaviors could be a manifestation of problems in the neural circuitry of vocal learning, in particular in the basal ganglia.
- The deficits in imitation witnessed across the whole autistic spectrum could be derived from affected vocal imitative ability.

Chapter five will be the conclusion of the whole dissertation. In the meanwhile, it will also sketch some questions needed to be addressed in future research.

2 The biological basis of vocal learning in animals including humans

The present chapter will introduce the biological basis, including behavioral, psychological, neuroanatomical, genetic, and evolutionary aspects of vocal learning in non-human animals as well as human beings. Then, based on current issues on vocal learning in nonhuman animals, several predictions and suggestions for further research will be proposed.

2.1 An introduction to the biological basis of vocal learning

Vocal learning is an ability to learn structurally organized sound patterns from conspecifics and modify vocalizations according to auditory feedback (Petkov & Jarvis, 2012). In the animal kingdom, the ability is prevalent in three groups of birds (songbirds, parrots, and hummingbirds), and some species of mammals (cetaceans, pinnipeds, bats, elephants) including humans.

2.1.1 Behavior

This section will introduce vocal learning in terms of behavioral aspects which include developmental analogy between birdsong learning and first language acquisition, sexual dimorphism and the multimodality of communication.

2.1.1.1 Vocal learning in development

Darwin (1871) had already implicated the analogy between birdsong and human language, and further speculated that language could have evolved from singing. Birds have been selected as great examples for studying vocal learning in various fields. Analogous to human speech development, birdsong learning contains three indispensable stages: A sensory (memorization) phase, when learners memorize the auditory input from conspecifics to set up a template; a sensorimotor phase, when juveniles produce deviated sound structures, normally called subsong, and concurrently modify them based on their auditory feedback; and a motor phase, when the sounds with complex structures are successfully acquired.

In the framework of generative linguistics, Chomsky has repetitively emphasized that first language acquisition is effortless. A neonate or even a prenatal baby with the functional commencement of the auditory system is exposed to the ambient language which enables her to master the language almost perfectly during the following years. When compared with second language learning, which starts at later stages of childhood period, first language acquisition is substantially less challenging. To draw an analogy, language acquisition is claimed to go through a similar process as that of birdsong learning in birds. In the normal hearing children, speech perception precedes speech production, which is similar to the memorization phase of birdsong learning. Babbling is a stage when babies utter just sounds but not speech, which resembles the subsong phase in birdsong learning. As the infant grows up, the first words, normally easily pronounced ones, are spoken. Utterances with structures like phrases and sentences will follow the n-word stage when only isolated words are produced. Both birdsong learning and first language acquisition are relatively limited in the timescale by the sensitive period (critical period), which is partly a reason why a second language is rarely learned as fluently as the first language. However, the cases of mastering a second language to a nativelike level are not rare, and surprisingly, quite a few species of vocal learning birds are able to change their song structures over the sensitive period, and such ability could even last for their whole life. Nonetheless, this analogy between second language learning and birdsong development has not been pointed out by others. The following section presents the proposal made on such analogy.

2.1.1.2 Sexual dimorphism

Compatible with Darwin's sexual selection theory, birdsong has been proposed to be an exclusive trait of male birds. Indeed, in most species of vocal learning birds, males are the vocal learners, whereas females are not. The variability and complexity of male birdsong attracts females for mating. However, recent studies have found that many species of female songbirds also sing, and female song exists in the common ancestor of modern songbirds (Odom et al., 2014). Such discovery poses questions to the classic Darwinian sexual selection account on the evolution of birdsong; however, it favorably supports the analogy between birdsong and human language since humans don’t show such a sexual dimorphism in terms of language. Moreover, even if the non-vocal learning females do not learn to produce songs, they still need to learn the songs auditorily to distinguish the male songs for evaluation purposes; therefore, in case of female birds, auditory learning sounds necessary too. The studies on female song learning also show that it takes less time for females than males to learn the same number of songs, and that auditory experience is necessary but not essential for males (Yamaguchi, 2001; Kriengwatana et al., 2016). This issue, together with the analogy between the second language learning and birdsong plasticity, will be readdressed in section 2.2.2.

2.1.1.3 Multimodality

At first glance, vocalization seems to be the paramount modality in communication among vocal learners. However, closer observations reveal that concurrent with the songs, movements of other body parts are also present in vocal learning birds. For example, beak movements, head motions and hops couple with the song during courtship display (William, 2001). Besides, a repertoire of song types has been shown to be coordinated temporally with a repertoire of dance-like movement (Dalziell & Peters, 2013). Furthermore, it has been found that birds exhibit deictic gestures (Pika & Bugnyar, 2011). All these studies signal the existence of a multimodal display in birdsong, both auditory-vocal and visual-wing/leg/beak modalities. This is in parallel with the existing proposal of multimodal origin of human speech. I argue that this multimodality of communication is essentially preserved along the evolution, even starting from invertebrates. I will flesh out my argument later in section 2.1.3. Additionally, more pieces of evidence suggest that such evolutionarily preserved multimodality does not merely exist in communication, but also in other cognitive domains. This will be shown in section 2.1.4.

2.1.1.4 Summary

In this section, the behavioral characteristics of vocal learning in vocal learning birds and humans were presented. As was mentioned, both song learning and language acquisition undergo three akin developmental stages, and song plasticity and second language learning seem to be closely analogous. The prevalence of female songs favors the analogy between birdsong and human language in terms of sexual dimorphism. Additionally, both birdsong and human language exhibit multimodality which refers to the co-occurrence of vocalization and movements of other body parts.

2.1.2 The role of memory in vocal learning

In this section, the well-known declarative/procedural memory model in psychology will be integrated with the vocal learning theory.

2.1.2.1 Declarative and procedural memory

Declarative memory is composed of facts and events that can be recalled consciously. It is assumed that the information in declarative memory could be explicitly stored and retrieved. On the other hand, procedural memory consists of skills that are unconsciously acquired, such as typing or riding a bicycle. With repetitive practice, we are no longer aware of the mastered motor production. Therefore, declarative memory is explicit, whereas procedural memory is implicit.

The declarative memory system is specialized for learning arbitrary bits of information and associating them. In this system, materials are learned rapidly and partially explicit, that means they are available to conscious awareness. The hippocampus, the entorhinal cortex and the perirhinal cortex in medial temporal lobe are responsible for learning and consolidating information in declarative memory. The procedural memory system is specialized for learning sequences and rules which are procedural in nature. Here learning is mostly unconscious and requires extended practice. . The system is also rooted in the network of cortical basal ganglia circuit. . Considering the similarities, the two memory systems have both competitive and cooperative interactions. While learning begins with declarative memory as it rapidly absorbs information, the procedural memory gradually learns analogous knowledge and eventually achieves the state of automaticity. Sequence learning however, requires an integration of explicit and implicit learning, in which both declarative and procedural memories are indispensable.

2.1.2.2 Birdsong structure

Birdsong exhibits a hierarchical structure consisting of multiple layers each of which is composed of discrete acoustic elements that are temporally ordered. In a sound spectrogram, from the lowest level, notes are combined into syllables, syllables are combined into motifs, and motifs are combined into sound bouts (figure 2). The sequence is in most cases fixed with only sporadic variation (Berwick et al., 2011). Linguistic terms---phonology and syntax---are also used in birdsong to describe the structure, in a way that phonology is employed to depict within-syllable structure and syntax to describe the arrangement of syllables into a large structure (Marler & Peter, 1988). Berwick et al. (2011) argue that both birdsong phonology and syntax are accessed by a finite-state automaton (FSA) which is much less complex than human language structure. The authors further argue that the fundamental difference between birdsong and human language is that birdsong lacks "compositional creativity", in the sense that the constituents of birdsong is unable to be recomposed into new meanings, even though the birdsong structures are complex enough. Recently, Suzuki et al. (2016) have reported that in Japanese tits, there is evidence for compositional syntax in calls (not songs,). However, we think that the experiment is not sufficient to prove compositionality as that in humans, rather, it is just a form of combinatory behavior which cannot reflect the compositional property of human language.

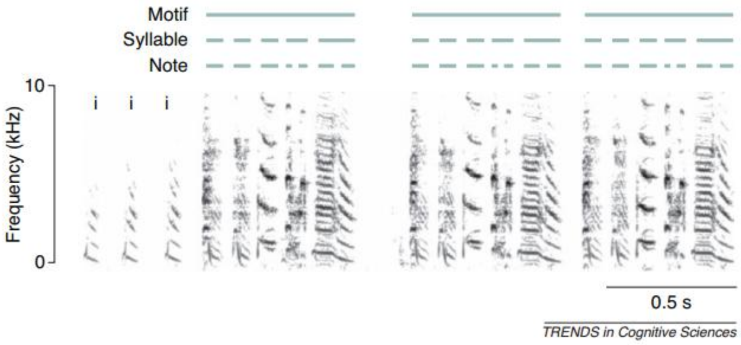


Figure 2: Sound spectrogram of a zebra finch with notes constituting syllables constituting motifs (Berwick et al., 2011)

2.1.2.2 Memory systems and vocal learning

Vocal learning is a type of sequence learning in auditory-vocal modality. Since both declarative and procedural memory play important roles in sequence learning, they should have a similar character in vocal learning. During the memorization phase, birds should use more declarative memory to remember the acoustic features of auditory input units like notes, motifs and phrases, and more procedural memory to learn the structure of the auditory template. In this process, the hippocampus—the neural basis of declarative memory is probably playing a more essential role. During the sensorimotor phase, birds should use more procedural memory to gradually grasp the correct sequence of sounds, which requires repetitive error-correction procedure. This reinforcement learning is analogous to any procedural learning of the sequences and rules in mammals that take advantage of the cortico-basal ganglia circuit like motor learning or habit formation. The way the volume of the hippocampus could be correlated to the open-endedness of songs in vocal learning birds will be elucidated in section 2.2.2.

2.1.2.3 Summary

This section focused on the combination of the declarative/procedural memory model in psychology with the vocal learning theory, stating the potential relationship between declarative memory and procedural memory and vocal learning. As a sequence learning in auditory-vocal modality, vocal learning process possibly involves both declarative and procedural memory.

2.1.3 Anatomy and neural correlates

The purpose of this section is to focus on the anatomy and neural pathways of vocal learning in the brains of vocal learning birds and humans, and also how analogous they can be at the neural level.

2.1.3.1 What is special about bird brain?

The pathways to vocal learning have been identified in three groups of vocal learning birds. Unlike mammalian brains, avian brains lack the six-layered neocortex and the uneven surface with gyri and sulci. Instead, they are composed of different nuclei and have relatively smooth surface. However, this does not prevent the birds from being one of the most intelligent creatures in nature. Studies have shown that quite a few of species of birds are as intelligent as chimpanzees. The results of activity-dependent gene expression and differential gene expression experiments provide evidence for the similar function between the mammalian cortex and avian pallium (Chakraborty & Jarvis, 2015). A recent report attributes this small-brain-high-intelligence achievement to the neuronal density. In other words, birds pack more quantity of neurons within the same volume of the brain area than mammals (Olkowicz et al., 2016). Moreover, the homology between avian brain and mammalian brain has been argued by Jarvis et al. (2005) that their cerebrums can both be subdivided into pallidal, striatal, and pallial areas (figure 3), the latter two of which contain the vocal learning regions, which will be discussed later (Jarvis, 2007). Therefore, comparing avian species and humans in terms of brain features seems reasonable.

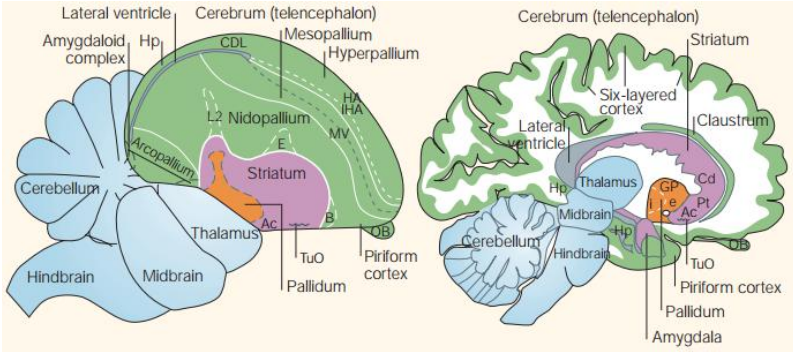


Figure 3: Modern consensus view of avian and mammalian brain relationships according to the conclusions of the Avian Brain Nomenclature (Jarvis et al., 2005).

2.1.3.2 Vocal learning pathways in birds and humans

Among the three vocal learning avian groups, songbirds are better studied than parrots and hummingbirds. Seven nuclei have been found in each group of birds specialized for vocal learning with four anterior ones, forming the anterior pathway and three posterior ones, forming the posterior pathway. The auditory pathway is not different in both vocal learning and vocal non-learning birds (Jarvis, 2009). In the following parts, it will be revealed that the auditory nuclei in vocal learning birds is likely to make a difference. Such connectivity responsible for vocal learning seems to be present in human brain in a remarkably similar fashion (figure 4). However, although it is widely acknowledged that humans are among the vocal learners, the neural circuit specific for vocal learning has not been confirmed in human beings.

AAC: central nucleus of the anterior arcropallium
ACM: caudal medial acropallium
Area X: areaX of the striatum
CM: caudal mesopallium
CMM: caudal medial mesopallium
DLM: medial nucleus of dorsolateral thalamus
DMM: magnocellular nucleus of the dorsomedial thalamus
DM: dorsal medial nucleus of the midbrain
HVC: a letter-based name
L1: field L1
L2: field L2
L3: field L3
MAN: magnocellular nucleus of anterior nidopallium
LMAN: lateral part of MAN
NLC: central nucleus of anterior nidopallium
RA: robust nucleus of the acropallium
VA: vocal nucleus of the acropallium
VLN: vocal nucleus of the lateral nidopallium

Table 1: Abbreviations of brain areas of birds.

2.1.3.2.1 The auditory pathway

The ascending auditory pathway and the descending feedback pathway are similar among vocal learning birds and non-vocal learning birds (see Jarvis (2007) for the detailed description of the pathway). In birds, the proposed function of the auditory pathway is to process complex sounds in a hierarchical manner (see figure 4): The acoustic features are processed by L2; the more complex acoustic aspects such as sequencing and discrimination are processed by L1, L3, and NCM; and the most complex aspects like finer sound discrimination are processed by CM (Jarvis, 2009).

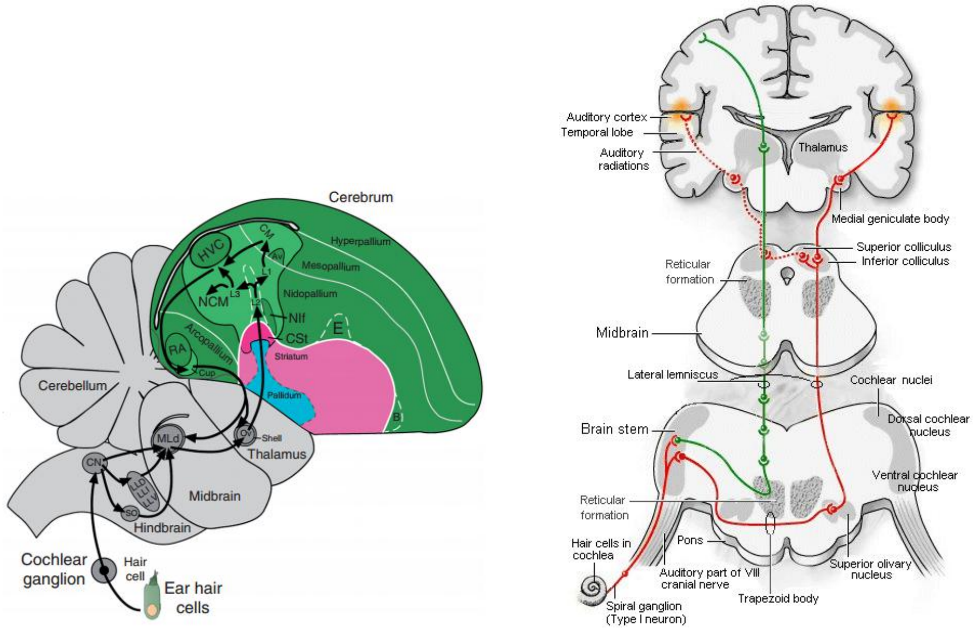


Figure 4: The auditory pathway in songbirds (left) and humans (right).

In humans, the auditory pathway starts from the cochlear nerve through all auditory structures that are organized tonotopically and hierarchically. Analogous to the functions of auditory areas in birds, the primary auditory cortex (Heschl´s gyrus) deals with temporally and spectrally simple sound stimuli, whereas the secondary auditory cortex (the supramarginal gyrus, superior temporal gyrus, insula and angular gyrus) and associative regions are in charge of processing complex sounds like speech. The associative regions send projections to the hippocampus which in turn projects back to the primary auditory cortex and associative regions. This reciprocal connection indicates that the hippocampus is probably participating in the auditory processing. Indeed, it has been reported that the hippocampus could inhibit redundant auditory inputs and detect novel auditory information (Kraus & Canlon, 2012).

Auditory input plays a crucial role in vocal learning process, as the auditory memory containing the structural template, the slots of which the sound units fill, is necessary for auditory feedback during the sensorimotor phase. The contribution of audition to language evolution will be shown in the next chapter.

2.1.3.2.2 The anterior pathway

In the brain of vocal learning birds, the four anterior nuclei constitute a loop comprising of a nucleus of pallium (MAN in songbirds, NAO in parrots) projecting to a nucleus of striatum (Area X in songbirds, MMSt in parrots), then to a nucleus of the dorsal thalamus (DLM in songbirds, DMM in parrots) and lastly back to pallial nucleus (Jarvis, 2007) (figure 5). This anterior loop is responsible for learning songs.

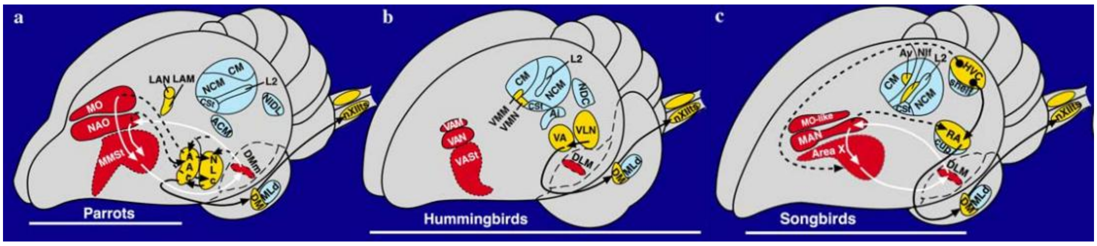


Figure 5: Proposed vocal and auditory brain areas among vocal learning birds on the left hemispheres. Red regions and white arrows indicate proposed anterior pathways, yellow regions and black arrows indicate proposed posterior pathways, dashed lines indicate connections between the two vocal pathways, and blue indicates auditory regions (Jarvis, 2009).

Anterior pathways have been proposed by researchers (e.g. Jarvis (2009); Fitch & Jarvis (2013)) to be analogous to the mammalian cortico-basal ganglia-thalamo-cortical loop, which has been shown to be responsible for habit formation, sequential learning, and reinforcing sequential learning process through trial and error correction (Graybiel, 2005). The loop starts from the premotor cortex, then projecting to medium spiny neurons of the striatum. The internal globus pallidus (GPi) in the striatum projects to the ventral lateral and ventral anterior nuclei of the dorsal thalamus. The projection then goes back to the premotor cortex where finally closes the loop. This basal ganglia circuit is most probably enhanced and specialized from general motor learning to vocal learning (see next subsection for the discussion of "motor theory of vocal learning origin" (Feenders et al., 2008)). Later studies indicate that the anterior vocal learning pathway is most probably starting from the inferior frontal gyrus where Broca's area (traditionally regarded to be responsible for speech production) is located, continues to the anterior striatum, then to the dorsal thalamus, and back to Broca's area (figure 6, b). Although the anterior pathway has not been confirmed in human brain, empirical studies identifying Broca-striatum-thalamic connections have emerged to favor the existence of the pathway. A diffusion-MRI tractography (DTI) study has detected connections among Broca's area, anterior putamen, medial globus pallidus, and ventral anterior thalamus (Ford et al., 2013). Teichmann et al., (2015) have also identified a Broca-caudate connection responsible for syntactic processing.

2.1.3.2.3 The posterior pathway

In the brain of vocal learning birds, the three posterior nuclei form the posterior pathway starting from a nucleus of nidopallium (HVC, NLC, VLN) to a nucleus of arcopallium (RA, AAC dorsal part, VA) to the vocal premotor neurons of midbrain (DM) and motor neurons of medulla (nXIIts) (Gahr, 2000; Jarvis, 2007) (figure 5). The nXIIts projects to the muscles of syrinx which is the vocal organ of birds. This posterior pathway is in charge of voluntary production of learned sounds.

The analogous posterior pathway in human brain is the direct cortico-laryngeal connection, which has been proposed to be the key neural connection underlying speech evolution in humans (Fitch, 2010). It projects from the laryngeal motor cortex (LMC) monosynaptically to the motoneurons of larynx located in the nucleus ambiguus. Such a projection from the motor cortex makes it possible to control the muscles of the larynx directly. This connection is responsible for voluntary sound production such as speech and song. In nonhuman primates, as believed before, such direct corticolaryngeal connection is absent, but an extremely weak connection between the motor cortex and the larynx in nonhuman primates has been found

(Arriaga et al., 2012), though it is too weak to rely on for a new function. The indirect connection with periaqueductal grey (PAG) in between is detected in all mammals for innate sound productions like laugh and cry. In this sense, among primates, only humans are equipped with two parallel connections for sound production—one for producing basic sounds and the other for complex learned sounds.

Consistent with the dual connection for sound production, some species of nonhuman primates possess two parallel—direct and indirect—connections for basic motor production and skilled motor production respectively (Simonyan, 2014). Considering that the two motor pathways resemble the two vocal pathways in terms of both structure and function, Simonyan (2014) proposes that the two parallel vocal pathways may undergo the similar path as the motor pathways in the evolution. Provided that it holds true, this will add another piece of evidence to support Feenders et al. (2008)'s "motor theory of vocal learning origin" in terms of the posterior pathway.

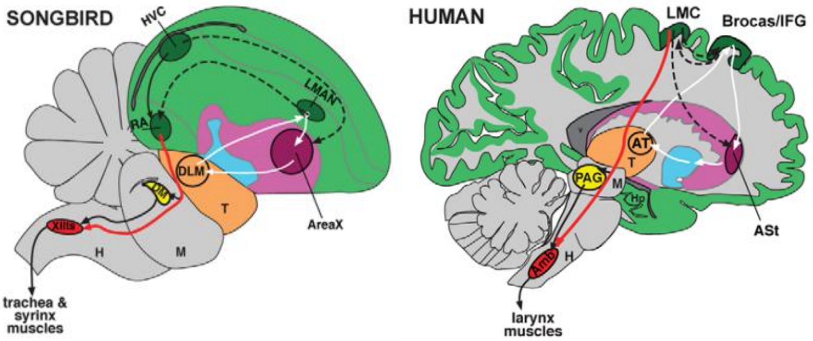


Figure 6: Red arrow: direct pallial-siryngeal connection (birds) and cortiolaryngeal connection (humans); white arrow: cortico (humans) (pallio (birds))-basal ganglia-thalamic-cortical (humans) (pallial (birds)) loop (Arriaga, Zhou & Jarvis, 2012).

2.1.3.2.4 The functions of the specialized nuclei

Further studies on the nuclei have revealed that during the sensorimotor phase, LMAN explores the motor output which results in more variability in vocal production. As the HVC takes over LMAN to project to RA, the produced song approaches stereotyped ones (Fee & Goldberg, 2011) (figure 7). This kind of reinforcement of vocal learning seems equivalent to motor learning and habit formation in mammals, of which the neural basis is the cortico-basal ganglia-thalamo-cortical loop (Groenewegen, 2003; Yin & Kownlton, 2006). The shift from LMAN to HVC indicates that during the subsong period, the production pathway for vocal learning is different from the period when birds start producing adult songs. Nonetheless, such a point has not been discussed in humans in an analogous way.

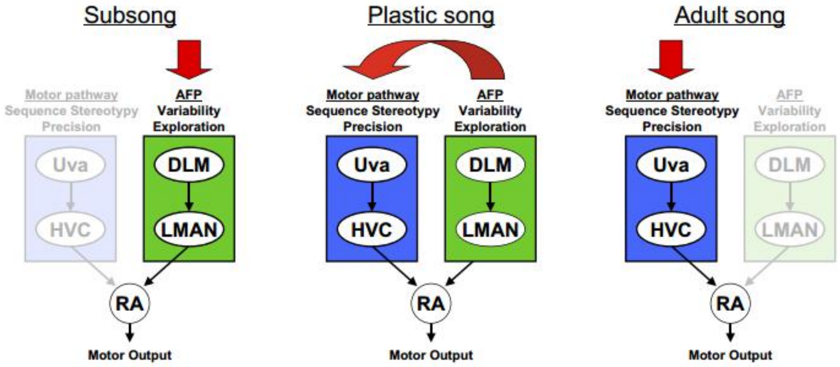


Figure 7: HVC is taking LMAN to project to RA for song production (Fee & Goldberg, 2011).

2.1.3.2.5 Left lateralization

Birdsong learning exhibits a left-hemispheric lateralization as is the case in human language learning and processing. On the basis of neuropsychological, neuroanatomical, and neurophysiological studies, it has long been accepted that language exhibits a left-hemispheric dominance. Left-sided dominance of neuronal activation in HVC is also found in zebra finches when they are learning songs, but the same activation in NCM is only present when they are exposed to tutor song rather than unfamiliar songs, suggesting that the lateralization affect is exclusive to the conspecific sounds and may be related to memory (Moorman et al., 2012).

2.1.3.2.6 Finding more in parrots

Parrots are one of the best vocal mimicry species. Vocal mimicry refers to the ability to mimic a heterospecific sound or even an inanimate sound. Recently, Jarvis and colleagues have identified the responsible pathway for vocal mimicry in parrots. They have found that different from other vocal learning birds, parrots have two sets of song systems: core and shell (Chakraborty et al., 2015). The core system is the same as in songbirds and hummingbirds, while the shell system that is unique to parrots, is perhaps fundamental to parrots' outstanding mimicry ability (Chakraborty et al., 2015) (figure 8). The core-shell-system discovery could be quite enlightening about the vocal mimicry ability in humans.

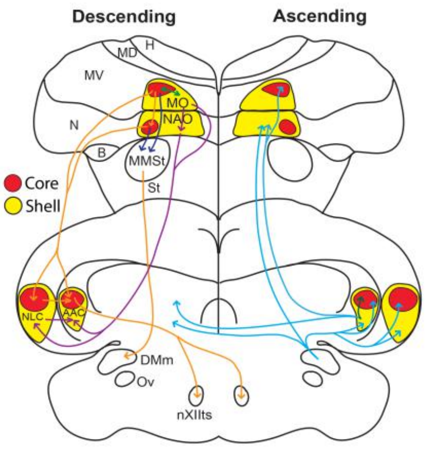


Figure 8: The core and shell system found in parrots (Chakraborty et al., 2015).

2.1.3.3 Summary

This section summarized the ongoing studies on the vocal learning pathways of birds, and the proposed vocal learning pathways in humans. It also illustrated the comparability between birds and humans from neural, anatomical, and functional aspects. Such comparison could shed light on the study of human vocal learning, and further guide the research on language and cognition.

2.1.4 Genetics

The molecular and genetic studies on vocal learning have proliferated in recent decades. With the discovery of the FOXP2 mutation in KE family (it will be fully discussed in the following section), the foxp family members have been under investigation in various species. The role of foxp2 in vocal learning will be discussed in detail in the following section. Dimerized with foxp subfamily members, not only foxp1marks the song nuclei in zebra finches (Teramitsu et al., 2004), but also is reported as a gene with alternation leading to language impairment in humans (Pariani et al., 2009). Cntnap2 is also discovered to be related to vocal learning. In the song production nuclei RA and LMAN of male zebra finches, cntnap2 transcripts are enriched, whereas in HVC this does not happen; and in Area X, its expression is reduced relative to the striatopallidum (Condro & White, 2014). In humans, a deletion of a single base pair in CNTNAP2 has been found in patients with cortical dysplasia-focal epilepsy (CDFE), which is characterized by language regression by the age of three (Strauss et al., 2006). Additionally, CNTNAP2 is related to autism spectrum disorder (ASD) (Bakkaloglu et al., 2008). Further studies reveal that CNTNAP2 is transcriptionally regulated by FOXP2 (Vernes et al., 2008). Moreover, cadherin is proposed as an important gene in language evolution by Matsunaga & Okanoya (2014). Cadherin expression has been found to be robust in RA in the Bengalese finch, changing from Cdh7-positive to Cdh6B-positive during the transition from sensory to sensorimotor learning phase (Matsunaga & Okanoya, 2014). The overexpression of Cdh7 affects vocal development in the Bengalese finch (Matsunaga & Okanoya, 2014).

More genetic and genomic studies have been carried out on vocal learning, but they are beyond the scope of this thesis. I believe research in this field will shed more light on the study of vocal learning in every aspect.

2.1.4 Evolutionary account

This section is focused on the evolutionary point of vocal learning. Here several influential proposals on the evolution of vocal learning from various respects will be briefly presented.

2.1.4.1 Sexual selection

It is inevitable to inquire how independent evolution of vocal learning diverse and distant lineages happens. There is a consensus among researchers that vocal learning trait is evolved independently in each lineage of birds and mammals. Sexual selection has been proposed as an evolutionary pressure to account for the successful selection of vocal learning because more pitch variations and complex structures are preferred by females. However, the variations of pitch and complexity of the songs attract predators' attention. This issue could to some degree explain why only sparse lineages of species with vocal learning ability have remained (Jarvis, 2009). Studies by Okanoya and colleagues on Bengalese finches have shown that in a more relaxed environment, the birds are able to produce more complex songs than their conspecifics in the wild (Honda & Okanoya, 1999), supporting Jarvis (2009)´s hypothesis. Nevertheless, as was stated in the previous section, a fair amount of females are also singers. This makes it more difficult to explore the evolutionary story on this trait. Supposedly, it could be sexual selection plus natural selection or other trajectories which need further exploration.

2.1.4.2 Motor origin of vocal learning pathway---Feenders et al. (2008)

In terms of the neural connectivity, Feenders et al. (2008) have proposed an elegant theory (as the authors call it a theory) on how vocal learning pathways have evolved, and it is called “motor theory of vocal learning origin”. On the basis of the result of IEG (immediate early gene) expression experiments, which show the genes that are activated when birds are hopping around, or are next to those activated when birds are singing (figure 9), the authors proposed a “motor theory of vocal learning origin” stating that vocal learning pathways have been duplicated from the existing motor pathways (figure 10). The authors speculated that it could be the case that one of the genes that are responsible for the motor pathway accidentally mutated, and this resulted in functional changes to a vocal modality. The authors predicted that this theory could also account for human vocal learning pathways, and other complex cognitive traits. Chakraborty & Jarvis (2015) further argue that the evolution of brain may be actually driven by pathway duplication, thus providing more theoretical supports for Feenders et al (2008)’s theory.

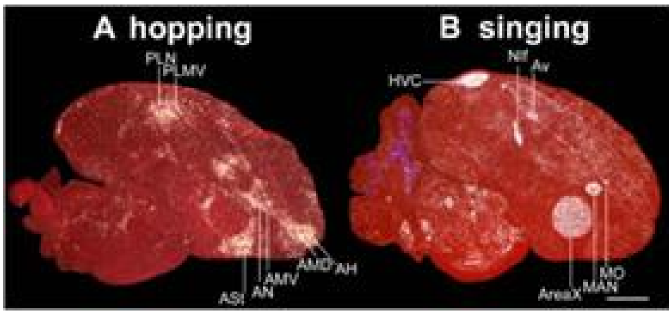


Figure 9: The IEG expression in zebra finches while singing (right) and hopping (left) (Feenders et al., 2008).

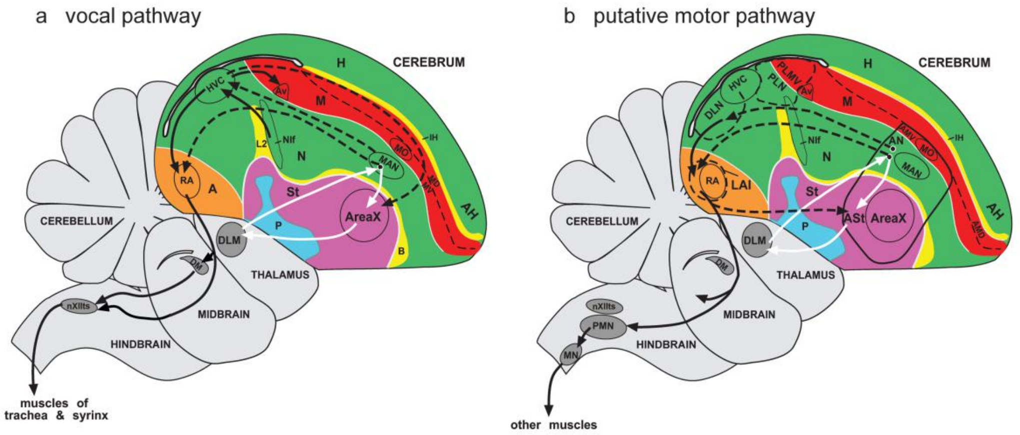


Figure 10: Vocal pathways and motor pathways: a, anterior and posterior vocal learning pathways in zebra finches; b, putative anterior and posterior motor pathways in zebra finches (Feenders et al., 2008).

2.1.4.3 Exaptation of the corticolaryngeal pathway—Fitch (2011)

In evolution, a shift in the function of a trait is described as exaptation. One well-known example is the feathers of birds. The feathers were primarily selected for insulation purposes, and later exapted for flight. This concept could be implemented across domains, ranging from a novel trait to a novel gene. Exaptation is an important concept to explain evolutionary phenomena and is in line with the "tinkerer" notion proposed by Jacob (1977).

In terms of the posterior vocal learning pathway, namely the direct corticolaryngeal connection, Fitch (2011) has hypothesized that the direct corticolaryngeal connection may have exapted from the corticospinal or corticobulbar connection, which is necessary for motor production in mammals. This seems partially in parallel with the “motor theory” proposed by Feenders et al. (2008) in vocal learning birds concerning movements. However, although Fitch states that the corticolarygeal connection is in the middle, with the corticobulbar connection above and the corticospinal connection below, he does not make it clear whether the pathway is an extension of the corticobulbar tract down to the spinal cord or a duplication of the corticospinal pathway, rather, he just addresses the issue from the functional perspective.. Nonetheless, anatomically speaking, it has been found that the direct corticolaryngeal connection is an extension of the corticobulbar tract to the nucleus ambiguus (Simonyan, 2014), which is supposed to be in parallel with the corticospinal tract.

2.1.4.4 Pathway competition—Deacon (1989/1992)

An interesting hypothesis has been put forward by Deacon (1989/1992) about the direct corticolarygneal connection, that there is a competition between the underlying neural pathway growth of innate calls and voluntary vocalizations. The neural system of voluntary sound production including speech and song in humans overrides the one of innate sound production like laugh and cry thanks to the enough space in human brains to allow the growth of the direct connection between cortex and larynx. Virtually, in evolution, there is an overall growth of hominin brain size, and over the course of human evolution, brain size more than tripled. Striedter (2005) called this Deacon´s rule, meaning ¨bigger equals better connection¨. To be specific, by virtue of more spacious room provided in human brain for the growth of new pathways, the corticobulbar connection got extended to the nucleus ambiguus, and at the same time got more strongly connected with the motoneurons of larynx. If such is true, it can be claimed that the factors that enlarge the brain size could play a role in this corticolaryngeal connection. The way this is connected with FOXP2 in humans will be explained in section 2.2.3.

2.1.4.5 Ackermann et al. (2014)

Ackermann et al. (2014) proposed subsequent evolution of the two vocal learning pathways in human brain, that are the monosynaptic refinement of corticolaryngeal projection subsequent with vocal-laryngeal elaboration of cortico-basal ganglia circuits driven by FOXP2 mutations. However, they have not mentioned how these two connections came into being. Although they presume the possible effect of the enlargement of human brain size in the evolution of speech pathway, they do not point out the specific mechanism behind this monosynaptic corticolaryngeal connection. As for the basal ganglia circuit, they attribute the two mutations of human FOXP2 to the morphological changes in basal ganglia, which are altered to the speech function in humans. In section 2.2.3, I will provide evidence that instead of both mutations serving to the basal ganglia morphological change, the two mutations of FOXP2 could play separate roles in the evolution of vocal learning pathways-- one for the posterior pathway and the other for the anterior pathway.

2.1.4.6 Genes and genomes

Pfenning et al. (2014) compared the sequenced genomes gathering from vocal learning species (songbirds, parrots, hummingbirds and humans) and vocal non-learning species (doves, quails and macaques), and generated a hierarchical genetic tree and a computational algorithm. The results confirmed not only the revision of the nomenclature and understanding of the relationships between avian and mammalian brains (Jarvis et al., 2005), but also the similarity between birdsong learning regions and human speech regions (figure 11). This makes the evolution of vocal learning an independent analogy as well as a deep homology.

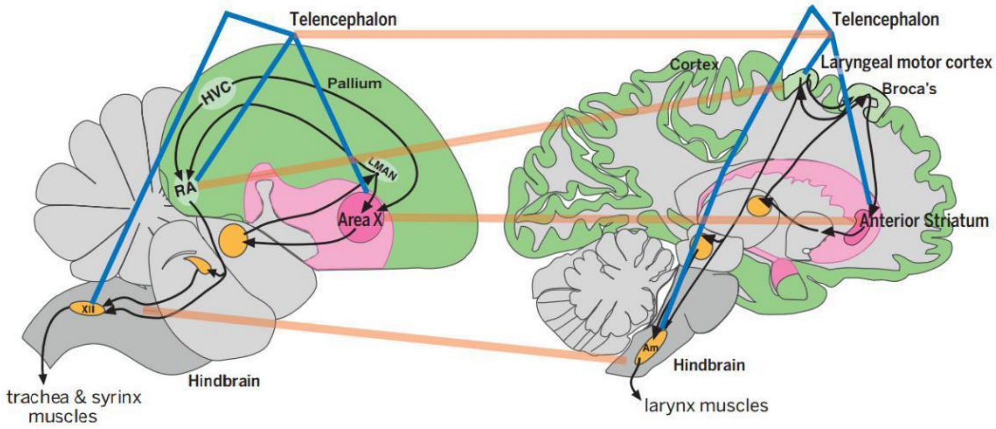


Figure 11: Molecular brain similarity between songbirds and humans (Pfenning et al., 2014).

2.1.4.7 Multimodal communication in nonhuman animals

Section 2.1.1.3 put forward the idea that movements are synchronized with birdsong which in turn proposes a multimodal exhibition in courtship display. This multimodality in communication also exists in other animals including humans. Although it has long been accepted that animals use various channels to communicate with conspecifics or heterospecifics, the studies on the integration of such multiple modalities are relatively few. Behavioral studies have shown that the multimodal communication is ubiquitous in animal kingdom. As an example, fruit flies (Drosophila) use visual, acoustic, olfactory, and tactile channels during courtship (Ewing, 1983). In early vertebrates-- i.e. fish, teleost fish (Bathygobius soporator) use visual, chemical, and acoustic signals by males in courtship (Tavolga, 1956). Studies on the cognition of fish have increased and become more informative during recent decades. For example, larval fish have been reported to use vocalization for communication (Bass, Gilland & Baker, 2008), and in coral reef fish, referential gestures for collaborative hunting have been found (Vail, Manica & Bshary, 2013). Furthermore, frogs have been claimed to use visual, audio, and tactile channels during courtship (Higham & Hebets, 2013).

Likewise, nonhuman primates also show simultaneous production of communication signals in vocal, facial and manual modalities. Although most of the studies on primates communication is unimodal (only 5% of a total number of 553 studies have examined multimodal communication in an integrated way (Slocombe, Waller & Liebal, 2011), it does not mean that the modality under the study is the only signal channel that primates emit. Multimodal signals can complete the intended information (Waller et al., 2013). In playful and aggressive contexts, apes use a slap gesture accompanied with facial expression to allow the conspecifics to respond properly (Rijksen, 1978). Gesture is visual, but can also be tactile or auditory (Liebal, Pika & Tomasello, 2004). Facial expressions like lip-smacking are visual and auditory at the same time (Micheletta et al., 2013). In bonobos, the ‘contest hoots’ calls are often directed at specific individuals and regularly combined with gestures and other body signals. , While the calls indicate the signaller´s intention to interact socially with important group members, the gestures add cues concerning the nature of the desired interaction (Genty et al., 2014). In data collected from 2,869 vocal events of 101 captive chimpanzees, approximately 50 percent were produced in conjunction with another communicative modality and about 68 percent were directed to a specific individual which is likely to include a signal from another communicative modality (Taglialatela et al., 2015).

2.1.4.7.1 The multimodal processing

In this section, thalamus, as a main role player in multimodal processing will be presented. Thalamus is a subcortical structure located in the very center of the brain of the vertebrates. It is a hub of information processing between different subcortical structures and cortex.

Multimodal information is processed along different sensory-motor pathways and integrated into unification. Such integration does not only occur in the higher association cortex, but already happened at low levels, even in subcortical areas (Tyll et al., 2011). The thalamus is an ancient subcortical structure which is preserved from early vertebrates. The anatomical connections with different sensory modalities (Cappe et al., 2009) and functional imaging suggest that the thalamus may play a crucial role in multimodal processing and integration (Tyll et al., 2011) (figure 12).

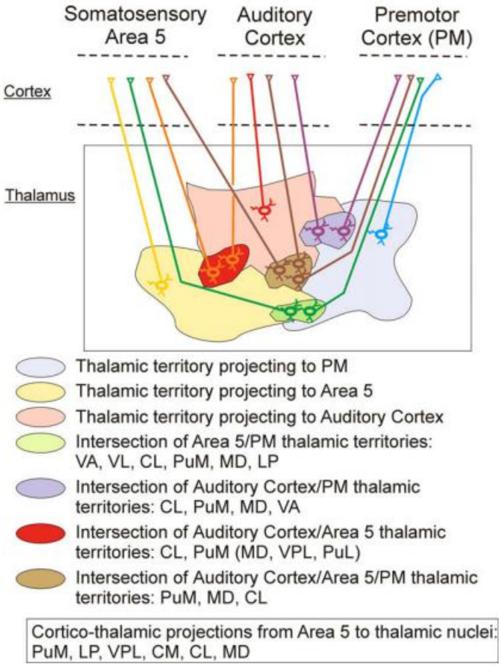

Somatosensory Area 5
Auditory Cortex
Premotor Cortex (PM)
Cortex
Thalamus
Thalamic territory projecting to PM
Thalamic territory projecting to Area 5
Thalamic territory projecting to Auditory Cortex
Intersection of Area 5/PM thalamic territories: VA, VL, CL, PuM, MD, LP
Intersection of Auditory Cortex/PM thalamic territories: CL, PuM, MD, VA
Intersection of Auditory Cortex/Area 5 thalamic territories: CL, PuM (MD, VPL, PuL)
Intersection of Auditory Cortex/Area 5/PM thalamic territories: PuM, MD, CL
Cortico-thalamic projections from Area 5 to thalamic nuclei: PuM, LP, VPL, CM, CL, MD

Figure 12: The interaction between the thalamus and the cortex for multisensory processing (Cappe et al., 2009).

This multimodal origin of communication signals described up to now is consistent with the proposed multimodal origin of human speech, which is the communicative signals in humans. I propose that foxp2 plays a generic role in the evolution of multimodal communication and across cognitive domains. This point will be explicated in the next subsection and chapter 3 will elaborate on how speech is evolved multimodally.

2.1.4.8 Summary

Here the existing proposals on the evolutionary account of vocal learning will be reviewed. Sexual selection might be the evolutionary pressure for the emergence of vocal learning trait, but it is an open question. The pathways of vocal learning have been proposed to be duplicated from the existing motor pathways (Feenders et al, 2008), which is consistent with Fitch (2011)'s hypothesis that the direct corticolaryngeal connection is exapted from corticospinal connection, and also in parallel with the idea of multimodal origin of speech. Besides, Deacon's competition hypothesis in which he emphasizes the point that the enlarged brain could be a causal factor for the duplication of motor pathways in vocal learning species including humans was depicted. Finally, it was mentioned that the multimodal communication signals in nonhuman animals is in parallel with the idea of multimodal evolution of speech in humans.

2.2 Predictions and suggestions

In this section, I will make several predictions for future research based on the studies on vocal learning reviewed above.

2.2.1 The subsequent evolution of vocal learning pathways

The theory of "motor origin of vocal learning pathway" (Feenders et al., 2008) do not say anything about the evolutionary order of the two vocal learning pathways, whether they are subsequently evolved or simultaneously co-evolved. I propose that it is more likely that the posterior pathway was evolved before and served as a prerequisite for the anterior pathway. In fact, Jarvis (2004) has already raised a similar argument that this direct connection (posterior pathway) in both birds and mammals "may be the only major change that is needed to *initiate* "(my emphasis) a vocal learning pathway. Only if the posterior pathway successfully grows at its full extent, the anterior pathway will be specified for auditory-vocal modality.

Empirical evidence indicates that it is very possible that it was the posterior pathway that was formed first, then lead to the formation of the anterior pathway. The posterior pathway is a direct connection between the motor cortex and the motor neurons of the larynx in the human brainstem. This direct connection vigorously exists only in vocal learners. In mice, such a projection from the motor cortex to the motoneurons of the larynx is also found, but it is not as robust as that in vocal learning species (Arriaga et al., 2012) (figure 13). As mentioned in the previous section, in the brain of nonhuman primates, which are traditionally regarded as vocal non-learners, the connection is not totally absent, and very weak projections have been detected (Arriaga et al., 2012). In suboscine birds that are not vocal learners and are a close relative to the oscine birds (vocal learners), Ai (intermediate acropallium) which is considered most comparable to the RA, has motor-related projections to midbrain/hindbrain, but not projecting to the brainstem (Liu et al., 2013) (figure 14). All these connections, rudimentary or weak, exist without any detection of the anterior pathway for vocal learning. Only the robust corticolaryngeal connection co-exists with the cortico-basal ganglia-thalamo-cortical loop for vocal learning. An early study by Devoogd et al. (1993) found that there is high correlation between residual volumes of HVC and area X. The authors suggested that “the evolutionary enlargement of one nucleus is associated with enlargement of the other”, further indicating that the evolution of the anterior nucleus is probably driven by the posterior nucleus. In addition, we never find a species with only anterior vocal learning pathway, rudimentary or weak, but without the posterior vocal production pathway.

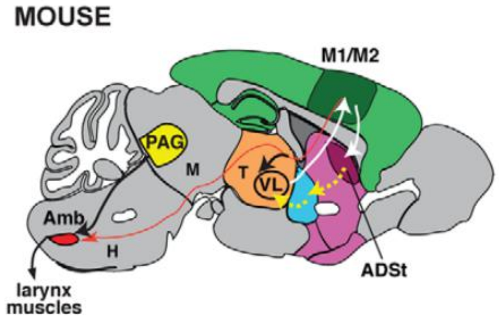


Figure 13: Mice brain with the direct corticolaygneal connection (Arriaga et al., 2012).

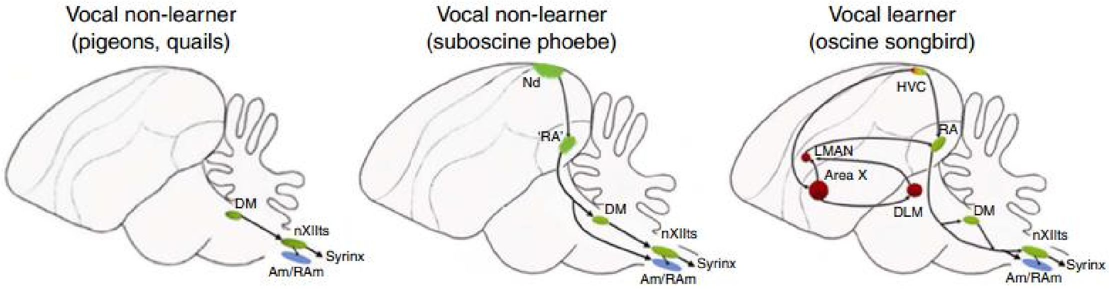

Vocal non-learner (pigeons, quails)
DM
nXIIts
Syrinx
Am/RAm
Vocal non-learner (suboscine phoebe)
Nd
'RA'
DM
nXIIts
Syrinx
Am/RAm
Vocal learner (oscine songbird)
HVC
RA
LMAN
Area X
DLM
DM
nXIIts
Syrinx
Am/RAm

Figure 14:Detection of the rudimentary posterior pathway in suboscine birds (Liu et al., 2013).

2.2.2 The role of the hippocampus in vocal learning

2.2.2.1. Brain size and vocal learning

It has been shown that there is a positive relationship between the brain size and cognitive abilities. The conception of brain size is not simply referring to the absolute volume of the brain, but rather to either a quotient of brain size relative to body size (encephalization), or to a quotient of an evolving brain area relative to a reserved brain area (relative brain size). Marler (2012) has speculated that "[t]he fact that the high end of avian encephalization is made up of orders that contain most of the bird species exhibiting vocal learning raises the possibility that vocal learning itself may be a factor in encephalization". Although Jarvis in his talks (e.g. Evolang 11) has explicitly mentioned that the size does not count but the network does for vocal learning, it is worth noting that songbirds and parrots indeed possess larger relative volume of striatum than other vocal non-learning birds.

With respect to the vocal learning mammals, the data also indicate a positive correlation between the relative brain size and vocal learning ability. Elephants, recently discovered as vocal learning species (Poole et al., 2005), have the largest relative cerebellum size of all mammals studied on average to date (Maseko et al 2012). Besides, they possess human-sized hippocampal formation (Stoeger & Manger, 2014). Odontocete cetaceans, and microchiropterans (a suborder of microbat) were also examined and found to have cerebellum clearly larger than the baseline determined from the analysis of primates (Maseko et al 2012). All this indicate that probably different species of vocal learners engage different brain structures in the process of vocal learning since the cerebellum has been reported to be involved in sound processing and production (Ackermann et al., 2007).

Coming back to birds, however, as for hummingbirds, even though their encephalization quotient is the biggest among avian species (Rehkamper et al., 1991), there is no sign of larger relative size of the striatum, though a significantly enlarged hippocampal formation has been reported (Ward et al., 2012). This enlarged hippocampal formation in hummingbirds is not surprising, because they are excellent avian foragers, that are capable of remembering the location and even the quality of each piece of food during their route, and such ability requires additional spatial-temporal information. Therefore, the observation that songbirds and parrots possess large relative size of striatum where vocal learning pathways are located implies that there could be a correlation between the relative brain size and vocal learning ability. Moreover, the fact of hummingbirds with enlarged hippocampus instead of bigger striatum suggests that the hippocampus may in some way contribute to vocal learning. The present thesis proposes that the hippocampus is not only involved in vocal learning process, but also plays a more important role in the song plasticity period, when vocal learner are still capable of acquiring new sounds as adults.

2.2.2.2. The role of the hippocampus in declarative and procedural learning

The hippocampus is essential for declarative memory (Squire, 1992; Tulving & Markowitsch, 1998), which includes both episodic and semantic memory (Tulving, 1972). In humans, episodic memory represents a person's experience in a temporal domain, whereas semantic memory collects concepts and knowledge we acquire. It has been argued that episodic memory is not specific to human beings rather, it is also present in nonhuman animals (Allen & Fortin, 2013). The hippocampus is sensitive to environmental novelty (events, places and stimuli) (VanElzakker et al., 2008). The multisensory information is presented to the hippocampus and processed through projections within substructures of hippocampal formation as well as between these structures and entorhinal cortex (Norman & O'Reilly, 2003). Memory is consolidated in the slow wave sleep (SWS) via information flow from the hippocampus to the neocortex (Rattenberg et al., 2011). Birds are the only species that possess SWS sleep among vertebrates other than mammals, which is necessary for the information transformation from the hippocampus to the cortex. This suggests that the brain waves in birds may be more similar to mammals than other nonmammalian species, further supporting the claim that birds may have episodic memory. Moreover, it might be interesting that hummingbirds demonstrate a special way of sleep called "torpor", during which the hippocampus could collaborate with the prefrontal cortex for memory consolidation. The bilateral damage to the hippocampi results in anterograde amnesia (the famous case of H.M.), but the long-term memories remain intact, again supporting the idea that memory consolidation requires the information transfer from the hippocampus to the prefrontal cortex.

Ample evidence has shown that the hippocampus is an essential subcortical structure involved in declarative learning. From the renowned case study of H.M., whose medial temporal lobe had been removed and had become amnesic but intact in remote memory, the underlying neural basis of declarative memory has been pinpointed to be the medial temporal lobe where the hippocampus is embedded. Furthermore, later neuroimaging studies on normal people support the critical role of the medial temporal lobe, especially the hippocampus in declarative memory (Eichenbaum, 2004).

Recently, literature has also provided evidence of hippocampus contribution in sequence learning (Davachi & DuBrow, 2015). Vocal learning is a type of sequence learning in auditory-vocal modality. Davachi & DuBrow (2015) stated that "repeated exposures to temporal regularities might drive the development and strengthening of a predictive code in the hippocampus that contains information about the order in which the sequence of items typically occurs". Functional MRI studies containing diverse sequence learning paradigms have also suggested that during sequence learning, the activation of the hippocampus is enhanced (Davachi & DuBrow, 2015). A case study of a patient with complete loss of the bilateral hippocampi and broader medial temporal lobe damage that had led to failure in simple sequential associations, also drives the attentions to the importance of hippocampus in sequence learning (Schapiro et al., 2014).

2.2.2.3 The possible involvement of the hippocampus in vocal learning

As an essential subcortical structure for learning and memory, hippocampus's impact in vocal learning seems highly probable. In spite of the fact that only male birds are vocal learners among avian species, female birds also need to memorize the conspecific songs in order to distinguish intraspecies from interspecies, and evaluate the body condition of males in breeding season. Thus, the memorization of the conspecific songs should occur in both sexes. In the IEG (immediate early gene) expression experiment on zebra finches, conspecific songs produced the highest densities of ZENK- and FOS-immunoreactivity across the NCM (caudomedialneostriatum), CMM (caudomedialmesopallium) and the hippocampus in both males and females (Bailey & Wade, 2005), which is a sign of possible involvement of the hippocampus in the auditory phase of vocal learning. There are claims that in contrast to mammals, whose hippocampi receive all sensory modalities with the assistance of surrounding areas of hippocampus like parahippocamlis (Allen & Fortin, 2013), the parahippocampus of birds only receives visual and olfactory inputs (Atoji & Wild, 2006). However, such claims exclude the possibility of the auditory information as a form of input to the hippocampus in birds. Not refuting the role of hippocampus in birdsong learning, the points mentioned may suggest a different route for auditory memory in birds. Furthermore, as far as we know, the studies which treat auditory information as an exceptional form of input to the hippocampus are only limited to chickens and pigeons, and the neural connections could be significantly different across species (e.g. the presence of vocal learning neural circuits in vocal learning species and the absence of them in vocal non-learning species), so the possibility of distinct neural connection in other avian species is not far-fetched.

Bolhuis & Gahr (2006) attempted to identify the neural substrate of the tutor song memory in vocal learning birds. According to the findings of a number of studies including Bailey & Wade (2005) that was mentioned above, the NCM and CMM located in the pallium of the birds are indicated to be “involved in processing of perceptual information concerning song complexity and in storage of song memory in songbirds and parrots”. This does not exclude the possible role of the hippocampus in auditory memory in birds, especially hummingbirds, as in mammals, neither the hippocampus nor the cortical structures are not the only brain structures which deal with sensory information (e.g. the temporal lobe is activated in processing auditory input as well). The NCM and CMM in birds have been proposed to be analogous to the auditory cortex in mammals (Bolhuis & Gahr, 2006), and they are most probably collaborating with the hippocampus in the same way the mammals do. The hippocampus and superior temporal cortex play a part in long-term auditory memory in humans (Teki et al., 2012). Furthermore, Kumar et al. (2014) have demonstrated that the hippocampus encodes the acoustic patterns that are learned implicitly. Thus, the hippocampus in birds may take part in the auditory processing in the memorization phase of vocal learning.

2.2.2.4. The relationship between the hippocampal volume and open-endedness of vocal learning birds

A positive correlation between the hippocampal size and open-ended vocal learning ability has been detected. Vocal learners were initially divided into two groups of open-ended learners and close-ended learners. An open-ended learner is capable of acquiring song plasticity as adults, whereas a close-ended learner cannot change their songs once the songs are crystallized. In other words, when a close-ended learner passes the critical period, their songs become stereotyped. Since critical period is varied in different species, a spectrum has been set up between the closed-ended species and the open-ended ones (figure 15). Coincidentally, this continuum is corresponding to the hippocampal size of the species, in a way that the most closed-ended species (e.g. zebra finch and Bengalese finch) maintain the smallest number, while the most open-ended species (e.g. starling) hold the largest number. Although the species in between are not the same, it has been proved that winter wrens with medium-sized hippocampus are actually able to modify their songs according to their neighbors (Camacho-Schlenker, Courvoisier & Aubin, 2011), thus the bird is claimed to be relatively open-end. Blackbirds with the biggest hippocampus along with with common starlingsare considered open-ended learners (Baptista & Gaunt, 1997). It has also been reported that canaries, common starlings, and red-winged blackbirds may acquire new songs over years or even throughout life (Baptista & Gaunt, 1997). The comparison suggests a positive correlation between the hippocampal volume and the open-endedness of the vocal learners, indicating that the hippocampus is probably the neural basis underlying the song plasticity of open-ended learners.

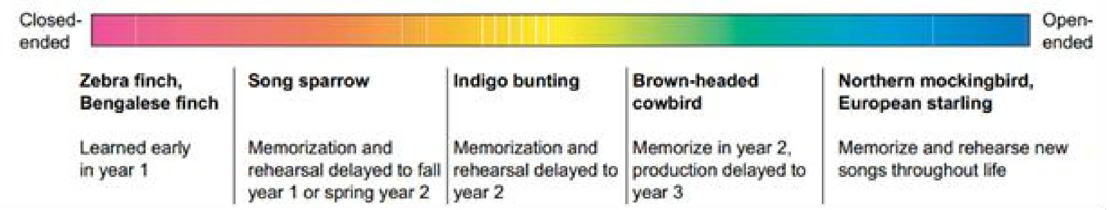

Closed-ended
Open-ended
Zebra finch, Bengalese finch
Learned early in year 1
Song sparrow
Memorization and rehearsal delayed to fall year 1 or spring year 2
Indigo bunting
Memorization and rehearsal delayed to year 2
Brown-headed cowbird
Memorize in year 2, production delayed to year 3
Northern mockingbird, European starling
Memorize and rehearse new songs throughout life

Figure 15: The spectrum of open-endedness of avian species (Brenowitz & Beecher, 2005).

Species		Log (hippocampus)
Melospizamelodia	song sparrow	1.315
Lonchurastriata	Bengalese finch	0.824
Taeniopygiaguttata	zebra finch	0.700
Troglodytes groglodytes	winter wren	1.093
Turdusmerula	blackbird (European)	1.593
Sturnus vulgaris	starling	1.459

Table 2: Hippocampal volume (the volumes are in cubic millimetres and have been log transformed) (Devoogd et al. 1993) with the hippocampal volume of song sparrow calculated by the author.

2.2.2.5 The role of the hippocampus in second language learning

The positive correlation between the hippocampal volume and the open-endedness of vocal learning could to a certain extent reflect how humans learn second languages as adults. Neural changes have also been captured in the cases of second language learning. In an intensive interpreter training program (Martensson et al., 2012) with three months of training, increasing volume of gray matter in the hippocampus, left inferior frontal gyrus, and superior temporal gyrus were detected in the trainees (Martensson et al., 2012). The left inferior frontal gyrus, including BA 44, 45 and 47, is considered Broca´s area in broad sense, and is proved to be involved in language processing, and the superior temporal gyrus plays an important role in auditory processing. The co-occurrence of increasing volume of the hippocampus and superior temporal gyrus is consistent with the simultaneous IEG expression in the hippocampus and NCM, and CMM in sensory vocal learning phase in zebra finches mentioned above. In addition, a recent paper demonstrates increased hippocampal activation in the initial stage of second language processing, but decreased activation after vocabulary consolidation (Bartolotti et al., 2016) This is a sign of the fact that the hippocampus is involved in both the sensory and sensorimotor phase of second language acquisition. Moreover, the hippocampal dentate gyrus is one of the regions in which neurons are generated throughout life (Drew et al., 2013), which provides more potentiality to the plasticity of the song or language learning in adulthood. Interestingly, LEF as a direct downstream target of FOXP2 (foxp2 as involved in vocal learning was discussed in section 2.1.4), regulates the generation of dentate gyrus cells, which relates the hippocampus to FOXP2, though FOXP2 has been reported not to be expressed in the hippocampus. Furthermore, a recent study for the first time showed that disruption of FOXP2 reduces the volume of subcortical structures including the hippocampus as well as the surrounded cortices in children, proposing a probable role for the hippocampus in the language impairment (Liégeois et al, 2016). The parallel indicates that the extended song learning in open-ended vocal learning birds could be analogous to second language learning in humans. Of course, such line of research requires further study in the future.

2.2.2.6 Summary

This section described the analogy between birdsong plasticity and human second language learning, and attributed the analogy to the hippocampus. Evidence of the possible involvement of the hippocampus in vocal learning in birds was provided. A positive relationship between the volume of the hippocampus and the open-endedness of diverse species of birds was found which suggests that the hippocampus can be influential in vocal learning, in particular song plasticity beyond sensitive period. The cases of the hippocampal growth of second language learning are also in favor of such analogy.

2.2.3 The role of foxp2 in the evolution of vocal learning pathways

The point proposed in this thesis is that in humans as in birds, the two vocal learning pathways undergo the same evolutionary route, with the posterior pathway evolving first, followed by the evolution of the anterior pathway. During this process, I hypothesize that the two mutations of FOXP2 in humans play separate roles in the formation of these two pathways: one (T303N) for the anterior pathway and the other (N325S) for the posterior one.

The most accountable gene concerning vocal learning and speech is foxp2. The study of KE family, a number of three-generation speech defected subjects due to the mutation of FOXP2 (R553H), provides enlightening clues to the possible responsible gene for speech (Lai et al., 2001). In evolution, foxp2 is a conserved transcription factor among vertebrates. Human FOXP2 experienced a >60-fold increase in substitution rate and incorporated two fixed amino acid changes in a broadly defined transcription suppression domain (Zhang et al. 2002). These two amino acid changes of the FOXP2 (N325S, T303N) occurred in the short timescale of human evolution split from the lineage of chimpanzees (Enard et al., 2002) (figure 16), pointing out that they may promote human specific trait(s) in the evolution. However, there has not been found any amino acid substitution of foxp2 that is shared between humans and vocal learning birds and mammals (Webb & Zhang, 2005) As such, the amino acid changes in humans may be pivotal in the exaptation of the neural circuitsand consequently give rise to new traits like speech in humans.

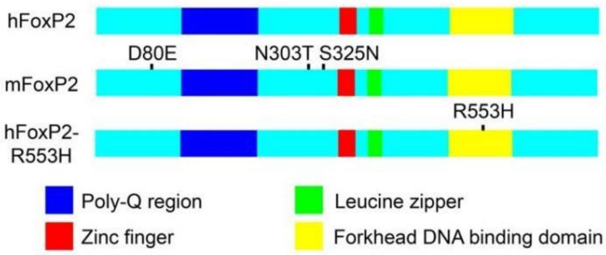


Figure 16: The two amino acid changes in human version of FOXP2.

In vocal learning birds, the knockdown of FoxP2 in Area X impacts the song learning of zebra finches, and results in an incomplete and inaccurate imitation of tutor song (Haesler et al., 2007). When mice are injected humanized version of FOXP2, there is a reduced dopamine level, increased dendritic length and long-term synaptic depression (Enard et al., 2009), and an accelerated relationship between declarative learning and procedural learning (Schreiweis et al., 2014), the neural basis of which are basically the medial temporal lobe and basal ganglia circuits. With this information in play, it seems that human version FOXP2 is getting involved in the basal ganglia circuit which is selected for an enhanced motor learning in vocal modality. I further argue thatonly one of the two amino acid substitutions (T303N) in humans have effects on this, whereas the other one (N325S) has impact on the posterior vocal learning pathway. Blending both mutations, only T302N mutation results in the same morphological change to the striatum in mice The findings show that the other mutation (N325S) does not take part in the reorganization of circuitry in the striatum, that is the anterior pathway (Bicanic et al., 2014).

Surprisingly, although these two mutations do not occur in other vocal learners, one of them (N325S) takes place independently in carnivores (Zhang et al., 2002) and a group of bat species (Li et al., 2007). Since "[s]everal studies have shown that phosphorylation of forkhead transcription factors can be an important mechanism mediating transcriptional regulation", "the human-specific change at position 325 creates a potential target site for phosphorylation by protein kinase C together with a minor change in predicted secondary structure", thus this mutation may have functional consequences (Enard et al., 2002). I hypothesize that this N325S may be crucial for the direct corticolaryngeal connection in humans. Overexpression of human, but not mouse, FOXP2 enhances the genesis of intermediate progenitors and neurons (Tsui et al., 2013). Usui et al. (2014) also suggest that by regulating the number of intermediate progenitors in the inner SVZ (subventricular zone) and outer SVZ, FOXP2 could be involved in volume growth of human brain.

If Deacon (1989, 1992) was on the right track, the formation of the posterior vocal learning pathway in vocal learning species would need more space than those of other vocal non-learning species. My suspicion is that there may be some sort of relationship between the genes responsible for big brain size and foxp2, particularly the N325S mutation. Investigating data from microcephaly, a neurodevelopmental disorder manifested as smaller head size, four genes have been identified as prime factors involved in disorders of neurogenesis–Microcephalin (MCPH1), ASPM (abnormal spinkle-like microcephaly-associated [Drosophila]), CDK5RAP2 (cyclin-dependent kinase 5 regulatory-associated protein 2) and CENPJ (centro-mere associated protein J). The mutation of these four cause a pathological reduction in human brain size (Bond & Woods, 2005). During the process of evolution, Microcephalin have regulated the brain size and have evolved under_strong positive selection in the human evolutionary lineage (Evan et al., 2005); human ASPM went through an episode of accelerated sequence evolution by positive Darwinian selection after the split of humans and chimpanzees (Zhang, 2003); Eveans et al. (2006) show that the protein evolution rate of CDK5RAP2 is significantly higher in primates than rodents or carnivores, and within primates it is particularly high in the human and chimpanzee terminal branches; Shi et al. (2013)'s study shows the hypo-methylation and comparatively high expression of CENPJ in the central nervous system of humans which suggest that a human-specific—and likely heritable—epigenetic modification have probably occurred during human evolution.

Apart from the four MCPH genes, beta-catenin has also been discovered to be correlated with brain size, to cadherins, and to some extent to foxp2. Beta-catenin (CTNNB1) signals are essential for the maintenance and proliferation of neuronal progenitors, controlling the size of the progenitor pool, and impinging on the decision of neuronal progenitors to proliferate or to differentiate (Zechner et al., 2003). On the other hand, Matsunaga & Okanoya (2014) focus on cadherins, and treat them as potential regulators in the faculty of language. Based on the recent findings on how cadherins are associated with various human psychiatric disorders, and the differential cadherin expressions between rodents and primates, they propose that novel brain functions could emerge by differential cadherin expressions. Containing redundancy and diversity at gene expression and function level, cadherin molecules may be good candidates to cause evolutionary changes in neural circuits subserving human language. Coincidentally, beta-catenin is an intercellular component of N-cadherin, and the Cdh2/catenin complex stabilizes synapse structure (Takeichi & Abe, 2005). In addition, a study (Rousso et al., 2012) on the neurogenesis of motor neurons (MNs) in the spinal cord identifies Foxp4 and Foxp2 as components of a gene regulatory network that balance the assembly and disassembly of adherens junctions (AJs) to promote neural progenitor cells (NPCs) proliferation and differentiation. The combined loss of Foxp2 and Foxp4 increase N-cadherin expression and retains NPCs in an undifferentiated, neuroepithelial state.

As stated by Scharff & Petri (2011), “identification of Foxp2 downstream target genes can help to pinpoint the cellular functions regulated by FoxP2 in a particular species, and comparing and contrasting FoxP2 targets in non-human animals with those in humans could provide important cues for potential functional changes occurred that might have contributed to the emergence of speech and language in the evolution". Huang et al. (2010) demonstrate that neuronal complexity controlled by p21-activated kinases (PAKs) is a key determinant for postnatal brain enlargement and synaptic properties. Interestingly, Pak3 is one of the FOXP2 direct targets (Vernes et al., 2011). Wang et al. (2015) have found that SLIT/ROBO, an important complex for axon guidance, is convergently downregulated in RA analogous area in three groups of vocal learning birds, dyslexia and other language disorders, and SLIT1 is a direct downstream target of human FOXP2. This connection provides insight to the involvement of FOXP2 in posterior pathway formation. As was already hypothesized, this could be attributed to one the amino acid changes in human FOXP2 (N325S).

In this section, I made a hypothesis that the two mutations of FOXP2 in humans play separate roles in the formation of the vocal learning anterior and posterior pathways. The T303N is responsible for the new specialization of the cortico-basal ganglia circuit, and the N325S is in charge of establishing connection between laryngeal motor cortex and the larynx in an indirect way which most probably play a role in the enlargement of human brain size. I suppose that in case of validity of my hypothesis, it can be predicted that there was a subsequent order between the two amino acid changes, that is T303N occurred before N325S. Evidence is lacking in current studies concerning this, and future research on the order of these amino acid changes will be telling.

2.2.4 The role of foxp2 in the multimodal evolution of communication

Comparative studies reveal that foxp2 can play a more generic role in the establishment of the neural scaffolding necessary for language acquisition and performance. In this section, I review recent studies on foxp2 and highlight the generic property of foxp2 that strikes us as critical in both evolution and development of the multimodal evolution of animal communication, as well as human language and general cognition.

2.2.4.1 Foxp in drosophila

As a homolog of foxp2, foxp has been studied among drosophila. The studies on drosophila provide us with the opportunity to investigate the role of foxp in an invertebrate model. Recent studies have shown that reduced Foxp in male drosophila disrupt pulse-song structure and sex-specific walking and flight (Lawton et al., 2014). This kind of higher locomotion control involves the brain structure called central complex (CX) in drosophila. The CX is implicated in courtship song production in both drosophila and grasshoppers (Popov et al., 2005; Heinrich et al., 2012). Foxp is strongly expressed in the CX and the protocerebral bridge (PB), which is a part of the CX. Interestingly, the CX in drosophila has been proposed to be comparable to the basal ganglia and the PB to the striatum in vertebrates, which is in accordance with the role of foxp2 expression in basal ganglia and striatum in vertebrates for sensorimotor learning and motor coordination. Besides, the roles of foxp in drosophila have also been explored in other domains. DasGupta et al. (2014)'s study demonstrate that it takes foxp mutants longer than the wild flies to make a perceptual decisions of similar accuracy, signaling that foxp is important for perceptual processing accuracy. Mendoza et al. (2014)'s study show that foxp in drosophila plays a crucial role in operant self-learning, "a form of motor learning sharing several conceptually analogous features with language acquisition", as well as habit formation which does not involve motor learning. Therefore, it appears that foxp is involved in multiple domains rather than just taking part in vocal production.

2.2.4.2 Foxp2 in mice

Mice and humans are similar in terms of neural expression and sequence of the encoded protein of FOXP2 (Campbell et al., 2009). This high degree of similarity has encouraged a bulk of Foxp2 studies on mice. Infant mice produce ultrasonic vocalizations (USVs) when they are isolated from their caregivers. It has been shown that the pups will produce abnormal USVs when the Foxp2 is knocked-out (Shu et al., 2005) and when the R552H mutation (a mutation similar to R553H in the KE family) is knocked-in (Fujita et al., 2008). Apart from the effects on vocalization, Foxp2 also has effects on neuronal development and formation of neural circuitry. The Foxp2 (R552H) mutation causes the immature development of Purkinje cells with poor dendrites in the cerebellum, which results in motor impairment in mice (Fujita et al., 2008). The heterozygous mutations of FoxP2 show significant deficits in species-typical motor-skilled learning, combined with abnormal synaptic plasticity in striatal and cerebellar neural circuits (Groszer et al., 2008). The heterozygous mutations also impair- sensorimotor association learning (Kurt, Fisher & Ehret, 2012), which shows that foxp2 may also be crucial for cross-modality association. Moreover, when blended with human version of FOXP2, the medium spiny neurons in the striatum exhibits increased dendrite lengths and synaptic plasticity (Enard et al., 2009).

2.2.4.3 Foxp2 in birdsong learning

The analogy between human language and birdsong was already presented in the previous section. Neurogenetic studies on birds have shed considerable light on human speech. Foxp2 turns out to be crucial for the formation of the anterior pathway (pallial-basal ganglia circuit, analogous to the cortico-basal ganglia circuit in mammals) and the vocal learning process. In zebra finches, foxp2 is predominantly expressed in the striatum, and there is a higher level of expression than the surrounding areas when vocal learning occurs (Haesler et al., 2004). The expression pattern of foxp2 changes seasonally and is corresponding to the social context. The expression of FoxP2 mRNA in Area X (the striatum) and the undirected song are negatively correlated (Chen et al., 2013), but this correlation does not exist in deaf birds (Teramitsu et al., 2010), therefore it shows that FoxP2 expression is modulated by motor activity and sensory activity (Wohlgenuth et al., 2014). This modulation effect in turn could be evident in the role of foxp2 in birds in sensorimotor association. In the process of song learning, the knockdown of Foxp2 in Area X in juvenile zebra finches affects the accuracy and completeness of the produced song (Haesler et al., 2007). It also deactivates the aforementioned socially contextual expression pattern, which is reflected by neural activity in LMAN, the downstream nuclei of Area X in songbirds' pallium (Wohlgenuth et al., 2014).

2.2.4.4 FOXP2 in multimodal problems of speech disorder

Inspired by the discovery of the FOXP2 R553H mutation in KE family, researchers investigate the role of FOXP2 in various language disorders. However, FOXP2 is unlikely to play any major role in the onset of autism or specific language impairment (SLI) (Newbury et al., 2002). Konopka et al. (2009) compared the human version of FOXP2 to the chimpanzee versionin order to explore whether or not they function differently. The authors uncovered genes that are differentially regulated upon mutation of the human two amino acids, some function of which is critical to the development of human central nervous system. Furthermore, the differential FOXP2 targets are involved in cerebellar motor function, craniofacial formation, and cartilage and connective tissue formation. This suggests that FOXP2 may be involved in establishing the neural circuitry and physical structures needed for spoken language. In this sense, FOXP2 targets are potentially linked to language impairment. As was mentioned in the previous section, one of the FOXP2 targets, CNTNAP2 has been tested associated with common forms of language impairment (Vernes et al., 2008), inclusive of SLI and autistic spectrum disorder.

It is worth noting that the symptoms of language disorders often discussed in the context of FOXP2 are not limited to the vocal modality. The speech-disordered children always suffer from general motor problems. Although the characteristics of childhood apraxia of speech (CAS) are manifested in the domain of phonological production at both the segmental and suprasegmental levels, the problems seem to originate from the more general motoric planning and programming (ASHA). Children who make inconsistent speech errors performed significantly worse than control on tasks requiring speech and dexterity of fine motor movement with time accounting as one performance factor, and children diagnosed as developmental verbal dyspraxia (DVD) had problem with the fine motor subset. "[T]his reflects deficits at the level of integrating sensory information into a plan of action and at the level of coordinating speech and dexterity of intricate movements" (Bradford & Dodd, 1998). Children with specific language impairment (SLI) also show motor problems. Specific structural motor anomalies were shown in SLI children parallel to their linguistic deficit (Roy et al., 2013). Substantial comorbidity exists between SLI and poor motor skill, implying that SLI is part of a broader phenomenon with motor incoordination as one element (Hill, 2001). SLI children also have difficulty in gestural comprehension and production compared with typical development (TD) children (Wray et al., 2015). Children with autism spectrum disorder (ASD) often exhibit language developmental problems. In tandem with speech deficits, ASD children also show gestural impairments and failure of integrating the gesture with speech production (So et al., 2014). Taken together, it seems that motoric (gestural) modality is also affected in speech and language disorders. In other words, if FOXP2 mutation affects speech development, and motoric (gestural) problem correlates with speech disorders, FOXP2 is most probably important for multimodal pairing.

2.2.4.5 Multimodality in other cognitive domains

Recent findings in the domain of music (obviously related to vocal learning) support the multi-modality discussion above. In most cases, music is accompanied by dance, and beat induction is an essential part of it. This is reminiscent of the coexistence of song and dance in avian courtship display. No matter when we sing or just listen to a small piece of music, we are perfectly capable of moving our body with the rhythm and make use of the beat induction, which is a cognitive mechanism that upon hearing a tone, we set up our internal beat, and use that internal representation to initiate our movement before any sound occurs at all, to synchronize/entrain with the music and/or with other dancers. This synchronization process requires fine auditory-motor coordination, an indispensable ability for vocal learning, which is the reason why Patel (2006) put forward his “vocal learning and rhythmic synchronization hypothesis”. We have mentioned that birds sing and pair with wing movement, and some birds have been shown to be very good dancers (e.g. Snowball). On the other hand, chimpanzees, imperfect vocal learners in the vocal learning continuum (Petkov & Jarvis, 2012), have been shown that their drumming behavior is a typical multimodal display that includes both vocal elements (pant-hoot) and swaggering and rushing about. According to Fitch (2015)’s definition, this should also be considered a form of dancing. A recent report by Dufour et al. (2015) also shows that a captive chimpanzee, called Barney, is able to do spontaneous drumming that has the properties found in musical drumming. Nonetheless, what a chimpanzee can do is far less than what a vocal learning bird can in terms of this synchronization. This in turn suggests that vocal learning could play an initial role in the vocal-motor entrainment.

Back to humans, parallels exist between music and speech in terms of multimodality. They follow a common timing process, suggesting that a cognitive rhythmic motor coordinator instigates such coordination (Mayberry & Jaques, 2000; Morley 2014). It has also been proposed that the co-occurrence of gesture and speech is associated with the prosodic rhythm rather than the lexical elements of speech (Trevarthen, 1999; Falk, 2004; Morley 2014). Furthermore, the affected KE family members are not deficient in either the perception or production of pitch, rather they fail in both perception and production of rhythm in both vocal and manual modalities (Alcock et al., 2000) This suggest that FOXP2 may underlie the rhythmic synchronization, which requires fine auditory-motor coordination as mentioned before. From a clinical perspective, Shi et al. (2015) illustrate the positive effects of musical therapy on non-fluent aphasia treatments. The data presented here make it clear that the mechanism is not specific to music or language but indeed domain general, and we speculate that foxp2 may play a crucial role in this.

2.2.4.6 Summary

In this section, by reviewing the role of foxp2 in multiple domains in both humans and nonhuman animals, I hypothesizes that foxp2 most probably play an essential role in the general mechanism that underlie multimodal communication and human cognition across domains. If this is on the right track, the multimodality casts doubts on the dichotomous hypotheses on either gestural or vocal (musical) origin of speech evolution, since gestures and vocalizations, as was described, seem to be closely associated with each other.

2.2.5 The role of breathing in the evolution of multimodal communication

In case of birds, although the wing movements are found to be synchronized with vocalization, close observation reveals that the wing movements are actually coordinated with breathing (Williams, 2001). It is known that vocalization is one form of behavioral breathing, in the sense that it requires specialized inspiration and expiration from basic respiratory rhythm. MacLarnon & Hewitt (2004) have pinpointed breathing as a human specific adaptation for speech, in the way that humans seem to master a better control of how to breathe, especially in speech production. In mammals, on the other hand, the vibration source of vocalization comes from lungs. Humans use intercostal muscles to hold subglottal air pressure, so that we can speak long enough before we take another breath. As MacLarnon & Hewitt (2004) have stated, the breath pattern of humans in vocalization is fundamentally different from that in nonhuman animals. Within one exhalation, we can speak a long sentence with multiple syllables, whereas in most nonhuman animals, one exhalation produces only one call. In this sense, the key role of breathing in speech evolution become obvious. Fitch (2009) has additionally taken into account the significant enlargement of the thoracic canal in modern humans (and Neanderthals) (MacLarnon & Hewitt, 2004) as a plausible fossil clue to human vocal control. The neurons feeding the intercostal muscles involved in breathing are located in thoracic spinal cord. This enlargement has provided modern humans with advantages for a better control of airflow during speech or singing. Since motor outputs are coordinated with breathing, including not only involuntary limb movements, but also volitional ones including gestures and vocal productions, I propose that the reason underlying the synchronization of gestures and vocalizations is actually a basic physiological one, namely breathing.

2.2.5.1 The role of the periaqueductal grey (PAG) in vocalization

The periaqueductal grey (PAG) has been assumed as an important element for innate vocalization. As an anatomic and functional interface between the forebrain and the lower brainstem, the periaqueductal gray (PAG) connects with diverse brainstem nuclei to coordinate specific patterns of cardiovascular, respiratory, motor, and pain modulatory responses. It is also engaged in processing fear and anxiety and producing vocalizations (Behbehani, 1995; Benarroch, 2012). The experiments of stimulating the PAG to produce vocalization can be dated back to Brown (1915), which were done on chimpanzee laughter. After that, stimulation of the PAG to produce vocalizations was carried on cats and monkeys (Magoun et al., 1973), rats (Waldbillig, 1975), guinea pigs (Martin, 1976), squirrel monkeys (Jurgens & Ploog, 1970; Kirzinger & Jurgens, 1991) and gibbons (Apfelbach, 1972). The more primitive sounds produced by animals that are associated with pain, fear or rage are mediated by multisensory information within the PAG, and the activation of such network is governed by activation of higher centers (Behbehani, 1995). Communicative sounds can also be produced by stimulating the PAG, but jointly with higher limbic or cortical areas, like anterior cingulate cortex or motor cortex. Lesions to the PAG may lead to mutism in humans (Esposito et al., 1999). The role of the PAG in vocalization has already been presented in early vertebrates. Kittelberger et al. (2006) and Kittelberger & Bass (2013) have demonstrated that the PAG robustly connects with a large number of vocal and auditory structures in a sound-producing teleost fish. This suggests that the PAG is a conserved gray matter phylogenetically responsible for social vocalization. Meanwhile, from the neural aspect, this is in line with the idea of the multimodal origin of communication.

2.2.5.2 The role of PAG in breathing

Subramanian and colleagues (Subramanian, Balnave & Holstege, 2008; Subramanian and Holstege, 2010; Subramanian, 2013; Holstege & Subramanian, 2015) have proposed that in mammals, it is periaqueductal gray (PAG) that is in charge of converting basic breath to behavioral breathing. Experiments on rat in vivo have demonstrated that stimulation of different parts of the PAG leads to increased or decreased activity of pre-l neurons, resulting in abnormal patterns of respiration (Subramanian & Holstege, 2013). Therefore, it can be concluded that the PAG plays an important role in controlling breathing rhythm. Using excitatory amino acid (d,l-homocysteic acid; DLH) to stimulate PAG, Subramanian (2013) found that the stimulation modulated both the late-l and post-l cells located in the medulla, that are proposed to be responsible for converting inspiration to expiration. The author further concluded that the PAG modulation is devoted to the conversion from eupnoea to behavioral breathing rhythm.

Given that PAG plays a role in both vocalization and breathing, and vocalizations and movements both are synchronized with breathing, it is possible that the underlying reason for the vocal-motor pairing lies in breathing, a very basic physiological activity of natural life. Besides, destruction of the direct corticolaryngeal connection doesn't generate mutism with intact innate vocalizations like laugh and cry, while destruction of the PAG connection renders mutism. This is another evidence to prove the importance of the PGA in vocalization. Consenting with Lund & Kolta (2006), Ackermann et al. (2014) attribute this PAG-derived mutism to breathing.

2.2.5.3 Foxp2 is expressed in the PAG

Admitting that the PAG greatly contribute to the multimodal origin of communication signal, and the prediction that foxp2 plays a generic role in multimodality, a correlation should be found between foxp2 expression and the PAG. In fact, foxp2 is expressed in PAG of mice (Campell et al., 2009). It is also hypothesized that the Kolliker-Fuse nucleus, a key structure for adaptive behaviors of respiratory network (Dutschmann et al., 2004), "can be uniquely defined in the neonate mouse by the coexpression of the transcription factor FoxP2 in Atoh 1-derived neurons of rhombomere 1" (Gray, 2008). Furthermore, Dutschmann et al. (2004) have also reported that "tauopathy induced progressive cell loss of foxp2-expressing neurons in the kolliker-fuse nuclei (KFn) is tightly linked to clinically relevant laryngeal dysfunction in tau-p301L mice" (Dutschmann, experimental biology 2016 meeting). More evidence is needed in the future research to test this hypothesis.

2.2.5.4 Summary

In this section, I further proposed that breathing plays an important role in the synchronization of multimodal communication. The vocal-motor pairing is probably simply biomechanical motor production synchronized with breathing. PAG--the neural basis of breathing could be involved in the vocal-motor synchronization, and also expresses foxp2 that I proposed in the previous section as the genetic basis of general mechanism of cognition.

3 The role of vocal learning in human language, communication and cognition

Language as a whole is a complex system involving diverse psychological functions (e.g. memory, attention etc.) and levels (e.g. phonology, semantics, and syntax), which seems to be unique to humans. However, this does not necessarily mean that the subcomponents that it presents are merely human features. Decomposing language into subcomponents enables us to realize whether there is any component that is genuinely unique to humans, or it is the system as a whole that is exclusive to humans. By rightly disentangling the language into its subcomponents and figuring out how they are organized into the language system, we will enlighten the study of human cognition as language, needless to say, is central in human cognition and communication. Moreover, we will eventually show that the majority of ingredients of language are also found in nonhuman animals. Vocal learning, the focus of the present thesis, illustrates this since it is a trait available in other nonhuman animals as we have demonstrated in the previous chapter. In this chapter, we will focus on the role of vocal learning in language and consequently in communication and cognition to some extent.

There is a general agreement that vocal learning abilities are necessary for speech (Bolhuis et al., 2014). However, what is speech in relation to language is much more debated. In the framework of generative linguistics, for instance, speech and sign constitute externalization. They are natural systems (in contrast to writing) coupled with the internal core language. In this case, externalization is secondary to the internal system integrated with the core linguistic system and the Conceptual-Intentional system (figure 16).

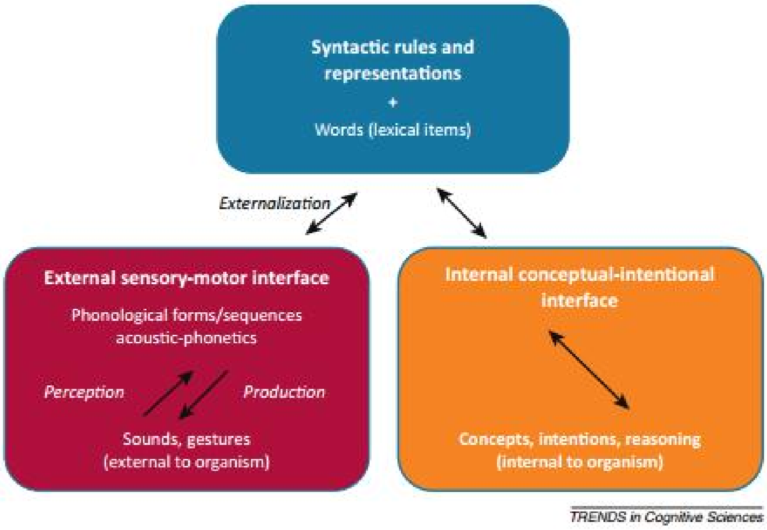

Syntactic rules and representations
+
Words (lexical items)
Externalization
External sensory-motor interface
Phonological forms/sequences acoustic-phonetics
Perception
Production
Sounds, gestures (external to organism)
Internal conceptual-intentional interface
Concepts, intentions, reasoning (internal to organism)
TRENDS in Cognitive Sciences

Figure 16: The architecture of language (*UG=universal grammar) (Berwick et al., 2013).

A qualification is however in order here. As shown in the above figure, externalization includes not only the motor part involved in speech/sign production, but also the sensory part for both modalities. In other words, externalization is a misleading label insofar it covers internalization as well. Bearing this in mind, the alleged ancillary nature of externalization with respect to the internal system does not seem so clear-cut: there would be an internal system that not only externalizes but internalizes. Such dichotomous internal/external division seems dubious, to say the least.

However, one could insist that the sensorimotor component, to use a more neutral term, is a necessary ancillary component, since there are two different sensorimotor modalities but only one language, that is human language. If this were true, vocal learning would only be relevant for the spoken modality and consequently not an indispensable factor regarding language. As will be argued later on, there is however a possibility that deserves consideration: sign could be an exaptation of speech that just materializes in a sign language, when the number of deaf people in a human group reaches a critical quantity. This view, which is supported by well-known and general facts about brain plasticity and explain the neural correlates of sign, does not preclude considering signs full-fledged languages. On the contrary, the fact that sign languages are like spoken languages is in line with our view. It is in a literal sense that we interpret assertions like the following one: signing is speaking with the hands and listening with the eyes. We will further insist on the plausibility of this view.

Therefore, assuming to be on the right track about the nature of sign, we hypothesize that vocal learning lays the foundational basis of language development (ontogeny) and evolution (phylogeny) for both modalities, with speech being primary. We will explicate our hypothesis from the following two main perspectives: 1) How sensorimotor integration in auditory-vocal modality has primacy in language evolution; 2) How sensorimotor integration in auditory-vocal modality has primacy in language development.

Additionally, we will expand our focus to deal with an inherent aspect of vocal learning which is imitation. First, there needs to be some background information on the issue. The importance of imitation in humans is considerable and goes beyond its role in language as suggested by the fact that Sapiens has been defined as Homo Imitans (Meltzoff & Moore, 1994). In this regard, it is not accidental that the influential paper by Hauser et al., (2002) published in *Science*, deals with this topic in three different (sub)sections. Knowingly enough, this paper introduced the distinction between the Faculty of Language in the Broad sense (FLB) and the Faculty of Language in the Narrow sense (FLN). The latter is defined as unique to humans, whereas FLB is made up of components present in other animal communication systems. FLN, according to the central hypothesis in the paper, would consist of recursion, which refers to the capacity to embed a phrase within a phrase.

As was mentioned before, imitation has been pointed out as a critical point in Hauser et al. (2002)'s study, and has been discussed rather extensively in three (sub)sections. First, when presenting "the comparative approach to language evolution", vocal imitation, as observed in birdsong, offers the best example of the relevance of analogies (or homoplasies) for the study of language evolution. Hauser et al. (2002) assert that the parallels between the ontogeny of speech and birdsong (critical period, babbling etc.) are 'intriguing'. Even insisting that the resemblances are a case of analogy and not homology, they recognize that there must be a common neural and developmental substrate for both speech and birdsong.

Next, in the subsection entitled *"How special is speech?",* it is mentioned that vocal imitation has been insufficiently studied, despite the fact that it "is obviously a necessary component of the human capacity to acquire a shared and arbitrary lexicon, which is itself central to the language capacity" (Hauser et al., 2002: 1574). Vocal imitation is, however, present in other animals like parrots, songbirds and cetaceans –with dolphins having a multimodal imitative capacity– although it is absent in nonhuman primates, which the authors find striking. The absence of visual imitation in monkeys is also stated in the subsection. All things considered, HCF see the study of the evolution of (vocal) imitation in primates so promising, but caution against the acceptance of finding the neural correlates for imitation as the last goal. Both imitation and intentionality are mechanisms for which it is necessary to know how they evolved instead of taking them for granted, and build human communication on top of them.

Finally, imitation reappears not in relation to communication but to cognition in the subsection *"The conceptual-intentional systems of non-human animals*". Here, after considering the big gap between the cognitive capacities of our closest relatives and their poor communicative abilities, the necessity of vocal imitation for lexical acquisition reenters the stage, now in form of the following dilemma: Should children’s capacity to acquire and recall words rely on a domain general mechanism or, in the light of the huge lexicons we have, should we think of an “independently evolved mechanisms”. At this point, the addition of the referential properties of words leads Hauser et al. (2002) to conclude that “many elementary properties of words”, if no precursor is found for them, should be reconsidered and then “this component of FLB (conceptual-intentional) is also uniquely human". Does “this component” refer to words or only to their referential properties? Do the “many elementary properties of words” include the imitation ingredient? That is not clear enough in the final paragraph of the section we are discussing. Be that as it may, there is a clear flaw in the whole argument which is putting words as ingredients of the C-I system only.

Words are by definition CI-SM pairs. Ironically, that vocal imitation reappears in this section stresses our point since this aspect of words falls under the SM dimension. It can be concluded that it is unfeasible to deal with words as though they belong to the C-I system alone.

All things considered, what Hauser et al. (2002) mention on imitation is inconclusive and seemingly wrong in the third instance. Words –we simply assume and will not expand on them–are cultural inventions built on top of sequences of sounds and generated by individuals who are endowed with a vocal learning capacity. This vocal learning capacity is at the base of our multimodal imitative skills and also of our non-vocal motor imitation (wrongly called visual imitation in the paper we are discussing). This is the hypothesis we want to deploy in this chapter where beside the arguments in favor of the primacy in phylogeny and ontogeny of the sensorimotor integration in the auditory-vocal modality, our continuist hypothesis will be confronted with the imitation theories that dominate psychological science in order to single out vocal imitation as the basic one. In that way, vocal learning will be additionally seen as a hugely significant promotor of the advancement of human general cognition.

3.1 Is speech special?

This is a question that have been raised and seriously discussed over the last century. It could be answered from two perspectives: a) Speech is special among other types of sounds b) Speech is special among other modalities. The former prompts research on the comparison between speech sound and other kinds of sounds, and the latter stimulates studies on the comparison between auditory-vocal modality and other modalities.

In this regard, Poeppel's (e.g. 2001) reply is more prominent than other positions and discussions. It says that speech is special in two different ways: it is special as a specialized auditory perception, i.e. speech sounds are not processed like other sounds, and it is hierarchically structured which goes beyond the extraction of statistical probabilities from the acoustic signal. We try to go further in this same line and propose that an enhancement of vocal learning might increase the structure building capacity that speech requires by combining sound and meaning.. In that way our proposal to some extent links Poeppel´s two separate points together. We will elaborate on our proposition of such a combination in the following sections.

Another point worth mentioning here is that the auditory-vocal modality *per se* is distinguished from other sensory-motor modalities, the source of which could date back to one of the Hockett's design features of human language (Hockett, 1960), i.e. total feedback. In Hockett's words,

"...the speaker of a language hears, through total feedback, everything of linguistic relevance to what he says himself. In contrast, the male stickleback does not see the colors of his own eye and belly that are crucial in stimulating the female. Feedback is important, since it makes possible the so-called internalization of communicative behavior that constitutes at least a major portion of 'thinking'".

The feature of total feedback in auditory-vocal modality is conspicuous but commonly neglected. To put it simply, what we hear *is* what we speak, and the auditory feedback helps adjust vocal production. In contrast, in visual modality, we need a further step of transformation from visual input to manual output. In other words, we utter sounds and hear sounds, whereas we see pictures but we paint, where pictures and painting are obviously of different necessity. This is also the reason that justifies Fodor (1983)'s action in taking light into account as one of his modules (or input systems), when was conceived that language acts as an input system.

3.1.1 Neuroanatomy of language

Before we explain our hypothesis in details, it is necessary to have a brief review of the studies on the neural basis of language. Studying the way language is implemented in the brain requires linking between mind and brain, that is to link the study of neuroscience and cognitive science. The endeavor to understand the neural basis of language could be traced back to 1800s, when Broca´s area and Wernicke's area regarded as the hallmarks of language production and comprehension were recorded, thanks to aphasic patients (Broca aphasia and Wernicke aphasia). Broca’s area was discovered by Paul Broca as a result of the discovery of a large lesion in the left inferior frontal gyrus in a patient who had serious problems with language production. Wernicke’s area was reported based on the observation of the lesion of patients who could produce fluent but nonsensical language and had impairment in language comprehension. Later, Wernicke's model was more clearly represented in

Lichtheim's model (Lichtheim, 1885).. Wernicke also predicted the incidence of conduction aphasia if the connection between Broca's area and Wernicke's area (the arcuate fasciculus) is damaged. The problem with the conduction aphasia is not in semantics, but in repeating utterance. Apart from the auditory and motor aspects (figure 16), Lichtheim incorporated regions for conceptual or semantic processing in his model. This was prompted with the advent of a new kind of aphasia, in which patients suffering from it had trouble expressing their thought, but could repeat normally. However, problems emerge afterwards, when the sole lesion to the areas does not suffice to the incidence of the corresponding descriptive aphasia, but other brain areas have to be appended, and reversely, one type of diagnosed aphasic patients also exhibit language problems typical of other types of aphasia, indicating that the aspects or subcomponents of language cannot be localized in the brain, but multiple areas and connections could be involved.

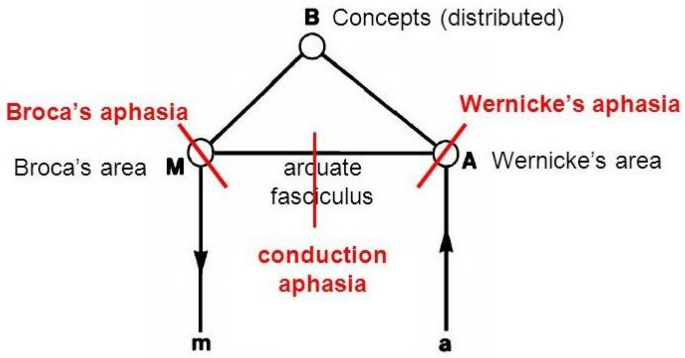


Figure 17: Wernicke-Lichtheim "House" Model.

As we gradually gain more knowledge of the brain and with the advancement of the techniques, we now are aware that language processing involves almost the whole brain including both cortical and subcortical structures. Mainly ethically limited, it is not easy to track the bundles in the human brain as it is in other nonhuman animals (of course it is not easy in some animals either). Moreover, neither is it easy to conduct a given linguistic task to be purely linguistic in the examination of the functional connections in the brain, as engaging other cognitive aspects such as attention and memory is inevitable. Thus, the identification of the neural connections exclusively for language is by no means an easy accomplishment if there were one. In spite of all difficulties, a large number of studies have been conducted on the language processing since decades ago, and more insights have been gained. In the next subsection, I will briefly review an influential hypothesis on the neural basis of language processing--a dorsal/ventral pathways hypothesis.

3.1.2 Dorsal/ventral pathways for language processing

This part will focus the dorsal/ventral pathways hypothesis for language processing. The hypothesis was first put forward in the domain of visual processing. This two-stream hypothesis has been well-established in the study of the processing of vision: the dorsal stream (where) is involved in the processing of the location of the object, and the ventral stream (what) is involved in the identification of the object (Milner & Coodale, 1995). Nonetheless, if we observe Lichtheim house model (figure 17) more carefully, we can realize that the reversal of the “roof” of the triangle remarkably resembles the current dorsal/ventral pathway model in language processing, being the sensorimotor dorsally located, and the semantic processing ventrally located. In the following subsections, Hickok & Poeppel and Friederici among others about the dorsal/ventral pathway hypothesis for language processing will be reviewed. In addition, we will review Van der Lely & Pinker (2014) and Boeckx (2016) that are supplemented to the current dual-stream hypothesis of language processing.

Hickok & Poeppel's hypothesis Inspired by the dual processing pathway for vision, Hickok & Poeppel (2004) made a similar hypothesis of functional anatomy of language processing The hypothesis follows as: A ventral stream is involved in mapping sound onto meaning, and a dorsal stream is involved in mapping sound onto articulatory-based representations, which coordinates the transformation between auditory representations of speech to motor representations of speech (figure 18A). The ventral stream projects ventro-laterally through the superior temporal sulcus (STS) towards the posterior inferior temporal lobe (pITL), including portions of the middle temporal gyrus (mTG) and inferior temporal gyrus (iTG) serving as an interface between sound-based representations in the bilateral superior temporal gyrus and widely distributed conceptual representations. The dorsal stream projects dorso-posteriorly through the posterior Sylvian fissure at the parietal-temporal boundary (area Spt) towards the frontal regions (figure 18B).

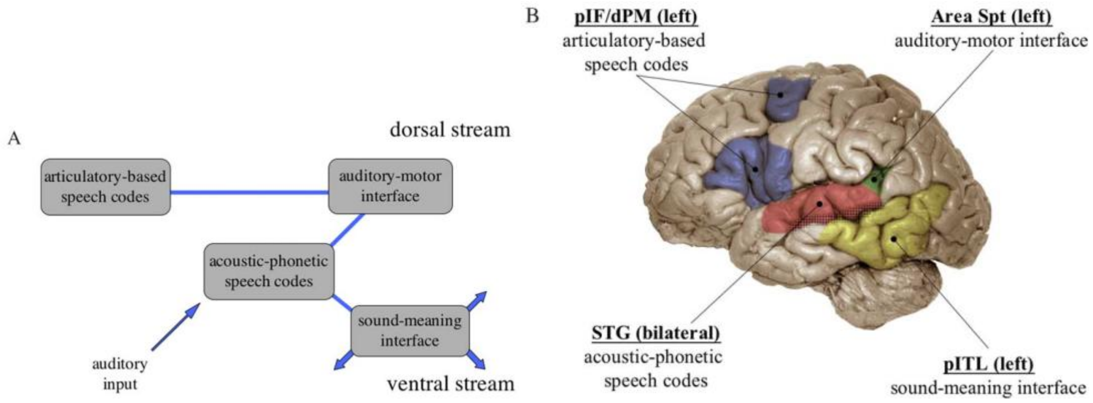


Figure 18: Dorsal and ventral streams for language processing (Hickok & Poeppel, 2004).

Nevertheless, although Hickok & Poeppel (2004) call their hypothesis the dual streams for *language* processing, they only touch on speech processing. At the level of brain, neither did they mention any subcortical structures possibly involved in language processing. The current consensus is that subcortical structures (the basal ganglia, the thalamus and the hippocampus along with the cerebellum) also have a contribution language processing. For example, Lieberman has provided ample evidence on the involvement of the basal ganglia in language processing as well as language evolution. In his paper, the reviewed findings identified circuits linking the basal ganglia, a subcortical structure dating back to early anurans, to various brain areas of both cortical and subcortical regions, which are active during linguistic tasks and high-level cognitive abilities (Lieberman, 2016). Besides, in the individuals with Huntington disease with early damage to the striatum, signs of speech problem (dysarthria) are also evident in the engagement of the basal ganglia in language processing.

Friederici's hypothesis Subsequently, closer observation of the dual pathways reveals that both the dorsal and ventral streams can be both structurally and functionally separated into two (or more) streams (figure 19). One of the dorsal pathways is a connection from the temporal cortex to the premotor cortex (BA6) via the inferior parietal cortex and parts of the superior longitudinal fasciculus (SLF) (Dorsal pathway I); the other pathway connects the temporal cortex to Brodmann Area 44 (part of Broca´s area) via the arcuate fasciculus (AF) (Dorsal pathway II) (Friederici, 2011). One of the ventral pathways connects anterior ventral inferior frontal cortex (BA45) along the temporal cortex to the occipital cortex via the extreme capsule fiber system mediating the inferior fronto-occipital fasciculus (Ventral pathway I); the other is a connection between the anterior and posterior regions within the temporal cortex (Ventral pathway II) (Friederici, 2011). The dorsal pathway I has been proposed to be involved in sound-to-motor mapping, and the dorsal pathway II is proposed to be responsible for processing complex sentences (Friederici, 2012), and it is not matured until the age of seven (Brauer et al., 2011). The ventral one is proposed to be responsible for building local syntactic structure.

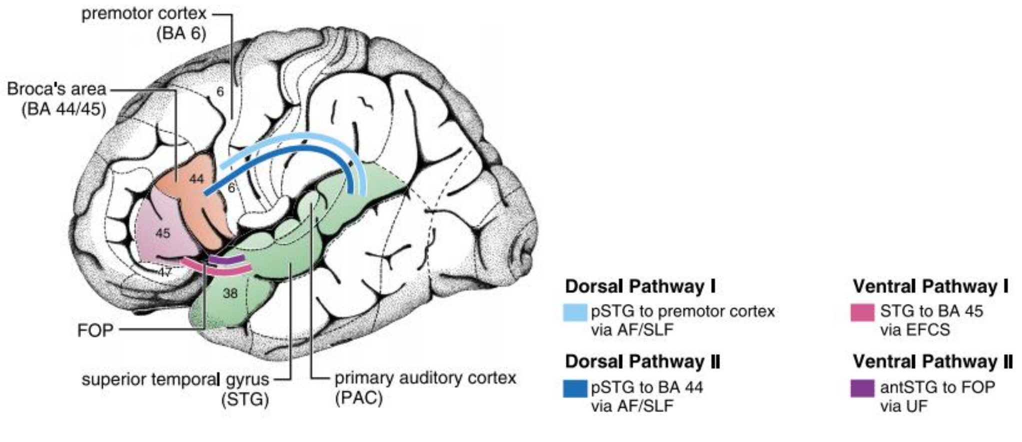


Figure 19: The subdivision of dorsal and ventral pathways for language processing (Friederici, 2011).

It is worth noting that the segregation of the dual pathway is not thoroughly from the start to the end, but occurs in the middle. In other words, given that the connection of temporal cortex and BA44 is postnatally mature (Brauer et al., 2011), we believe it is possible that the dorsal pathway in charge of complex sentence processing most probably diverged from the one for auditory-motor integration in both development and evolution. If this is true, the pathway for complex sentence processing has been generated from the pathway for auditory-vocal integration, which is a key neural connection for vocal learning. Moreover, literature has suggested that the dorsal pathway is taking part in both phonological processing (Murakami et al., 2015; Schwartz et al., 2012) and syntactic processing (Goucha et al., 2017), suggesting that phonology and syntax may have come from a common neurological origin, which in our proposal is the under-generated dorsal pathway in children. Still, Friederici and colleagues only focus on cortical connections without mentioning any subcortical structures or connections that could be engaged in language processing, further suggesting an incomplete picture of language processing for the dual-pathway hypothesis.

Van der Lely & Pinker (2014) Integrating cortical and subcortical regions, Van der Lely and Pinker (2014) attempted to investigate the neurological basis of language processing using participants with specific language impairment. Consistent with Friederici's hypothesis, the results of their experiment have shown that the dorsal pathway is involved in the extended syntactic computation, which is nonlocal, hierarchical, abstract, and composed, while the ventral pathway is involved in the basic syntax, which is local, linear, semantic, and holistic. What is worth mentioning is that the authors have additionally proposed that subcortical structures--the basal ganglia and the hippocampus--may also be involved in extended and basic syntax. In this sense, Van der Lely & Pinker's study parallels Ullman's proposal of procedural/declarative framework on language learning. It was mentioned in the previous chapter that the cortico-basal ganglia circuit is engaged in the procedural learning which is implicit and rule-based, whereas the medial temporal cortex particularly the hippocampus is engaged in the declarative learning which is explicit and straightforward.

Interestingly, this procedural/declarative learning hypothesis in general has been attempted to be tested in nonhuman animals regarding foxp2. Injected with human version of FOXP2 in mice, the link between the declarative and procedural learning is accelerated (Schreiweis et al., 2014). In the previous chapter, we raised the point that the function of foxp2 could play a domain general role in cognition. It is conceivable that the human version of FOXP2, enhancing the transition between the declarative and procedural learning, subserves language learning.

In modern linguistic theory (the minimalist program), the hierarchical structure building ability, which is dubbed as *merge* by Chomsky, is regarded as the core computational tool of language faculty. The hierarchical structures are manifested as *recursion* which is regarded as a defining property of human language faculty. If our hypothesis that vocal learning plays an essential role in language development and evolution is on the right track, the ability of recursive structure building would be derived from vocal learning ability.

Boeckx (2016) has recently put forward an interesting proposal from the perspective of Darwin's descent with some modification. By conceptualizing recursion as the capacity for sequencing sequences, he combines his proposal of globularity and recursion. Mainly in agreement with the dorsal/ventral language pathway proposal and presenting the evidence of the invasion of the parietal lobe in-between frontal and temporal lobes, which makes an indirect dorsal fronto-parieto-temporal connection, he hypothesizes that the dorsal dimension for two-dimensional tree-like hierarchy is the result of pairing two evolutionary existing networks for one-dimensional finite state sequences, namely fronto-parietal and fronto-temporal respectively. Boeckx argued that since these two evolutionarily ancient networks are both involved in finite-state computation, with the fronto-parietal network envisaged by Dehaene et al. (1998) as a sequence producer, and the fronto-temporal network already presented in primate audition, the pairing of both will render a two-dimensional computation, yielding the recursive representations.

Although in his paper Boeckx explicitly admits that both cortical and subcortical structures contribute to the neurobiological infrastructure of language, he only emphasizes the role of parietal lobe, downgrading the role of other subcortical structureslike the basal ganglia. We find Boeckx's view on sequences of sequences interesting, but we think it is more conceivable that the neural substrates for vocal learning are the best candidate for the sequential computation, at least for one of the sequences of sequences. Meanwhile, we believe the two mutations of human version

FOXP2 play a key role in the enhancement of declarative/procedural learning subserving language acquisition, and POUF2 plays a crucial role in the enhancement of hierarchical structure building ability. We will explain our point of view in details in the next section.

3.1.3 Cortical sequences plus cortico-subcortical sequences

What Boeckx (2016) has explained as the core computation of human language, recursion could be understood as a pairing of sequences with sequences. In contrast to Boeckx who presumed this sequences with sequences at the cortical level, in this section, we propose that this pairing is most probably composed of both cortical network and subcortical network. To be more specific, one of the sequences comes from dorsal pathway for sensorimotor integration, whereas the other ones originate from the cortico-basal ganglia circuit. The cortico-basal ganglia pathway for vocal learning will connect the arcuate fasciculus for auditory-vocal integration, then the learned sounds will be vocalized via the corticolaryngeal connection. As we have stated in the previous section that dorsal pathway II is postnatally generated, it could be the case that the enhancement of vocal learning in humans prompts the segregation of the dorsal pathway into dorsal pathway I and dorsal pathway II. The dorsal pathway II in human brain has been proposed as the crucial connection for complex syntactic processing (Friederici, 2011). Moreover, we will show in the next section that humans possess more massive and stronger white matter fibers of the arcuate fasciculus and superior longitudinal fasciculus compared with nonhuman primates, which could accelerate information transmission faster, thanks to the two amino acid changes of FOXP2. Furthermore, we also stated that human version of FOXP2, when blended in mice, makes medium spiny neurons in the striatum exhibit increased dendrite lengths and synaptic plasticity (Enard et al., 2009). Therefore, it is conceivable that the sequences of sequences could have come from the connection between the dorsal pathway and the cortical-basal ganglia circuit.

For such sequences of sequences, we propose that one sequencing is run by the cortical-basal ganglia circuit, and the other one by the dorsal pathway (white matter fibers connecting Broca's area and Wernicke´s area, namely the arcuate fasciculus and superior longitudinal fasciculus). On the one hand, it is established that the cortico-basal ganglia circuit is responsible for sequential learning, which provides input to the motor cortex for production via sensorimotor transition circuit. Thus, it is viable to presume that the anterior vocal learning pathway could offer one of the sequences required for recursive computation. On the other hand, based on assumption as well as genomic evidence (Pfenning et al., 2014), it can be claimed that in songbirds, Broca’s area is homologous to the LMAN, and Wernicke’s area is homologous to the HVC. Therefore, the connection between the LMAN and the HVC in songbirds could be functionally analogous to the dorsal pathway in humans. Evidence reveals that this LMAN-HVC connection is involved in song structure variability and sensorimotor transformation, when the juveniles are learning the song structures (Hamaguchi & Mooney, 2012). When the connection is transferred from HVC to RA, the produced songs will be stereotyped. As we mentioned in the previous chapter, RA projecting to the nXIIs in the brain of songbirds is analogous to LMC projecting to the brainstem in human brain (figure 20). As will be argued in section 3.2.1.2, in humans, the arcuate fasciculus is more massively and strongly connected than that in nonhuman primates. Additionally, recently in their paper, Goucha, Zaccarella & Friederici (2017) provide neurological evidence, emphasizing the role of "arcuate fascicle responsible for the rule-based combinatorial system, implementing labeling and giving rise to hierarchical structures". All these point to the possibility of the dorsal pathway being the other sequences of sequences of sequences.

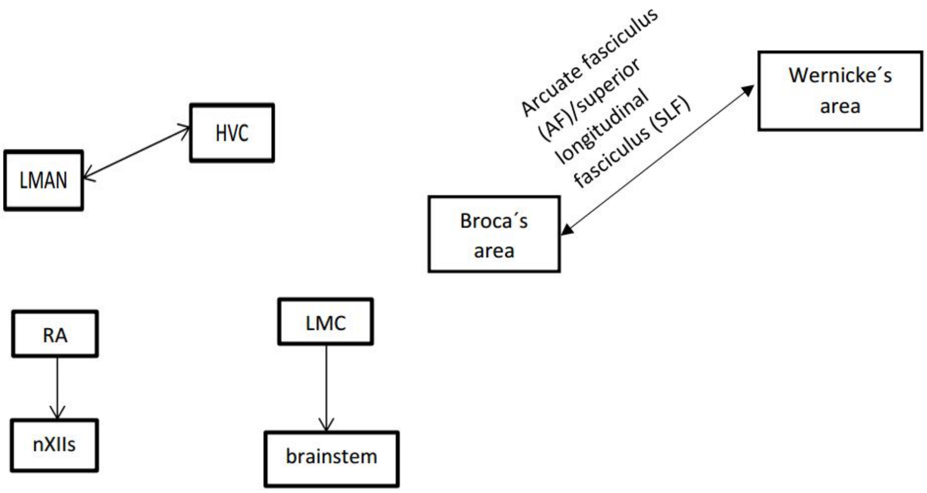

LMAN
HVC
RA
nXIIs
LMC
brainstem
Broca´s area
Wernicke´s area
Arcuate fasciculus (AF)/superior longitudinal fasciculus (SLF)

Figure 20: The analogous connection for sensorimotor integration and the posterior pathway in birds and humans.

Therefore, the following question needs to be addressed: Why cannot vocal learning birds (and perhaps other vocal learning species) acquire language ability? Firstly, we will have a look at different functions of the brain structures. The stronger and more massive connection between the superior temporal lobe with the parietal lobe, frontal lobe and the striatum presented in humans is not available in birds. Although connection between the LMAN and the HVC is engaged in the sensorimotor transformation, the connection is not strong enough as found in human arcuate fasciculus. In addition, although homolog has been found between the HVC and Wernicke's area, the advanced auditory ability is not located in the HVC (but rather in ACM and CMM), yet it is located in the temporal lobe in humans. We emphasize here that sequential learning could be cross domain, not necessarily in auditory-vocal modality (see next section of the discussion of action grammar). The second reason lies in the absence of words (or meaning) in birdsong or other communicative signals. Birds indeed have developed culture-like community to some degree, but not complex enough to prompt words and meanings.

3.1.3.1 What FOXP2 and POU3F2 could have contributed to the sequences of sequences

Thanks to the discovery of KE family, FOXP2 has been commonly accepted as one of the crucial genes that is involved in language. However, what we want to emphasize here is that the function of foxp2 is better understood as a domain general gene, rather than a language gene or the language gene. Yet, this does not mean that foxp2 does not play a role in language development and evolution. The enhanced sequential learning found in human language could probably be a by-product of the enhanced function of human version of FOXP2. Furthermore, we made it clear in the previous chapter that language has evolved on the basis of advancement of multiple cognitive abilities, so FOXP2 in humans that is underpinning domain general cognition has played a role in language evolution.

When it comes to sequence or sequence learning, the cortico-basal ganglia circuit is widely discussed. The cortico-basal ganglia circuit underlies motor sequence learning, and we have stated that song learning in vocal learning birds considerably engages the cortico-basal ganglia circuit as well. This should be one of the language ready abilities across domains. Referring back to foxp2, human version of FOXP2 has been studied in mice. Evidence has shown that when blended with human version of FOXP2, the synaptic plasticity and dendrite connectivity are increased in the cortico-basal ganglia circuits (Enard et al., 2009). This morphological change may have given rise to a better information processing ability. Indeed, as we mentioned in the previous chapter, the humanized FOXP2 in mice accelerates the transition between declarative learning and procedural learning, and the mice can quickly integrate the visual and tactile clues (Schreiweis et al., 2014), suggesting that human version of FOXP2 may stimulate higher learning speed as well as multimodal processing speed. If language development and evolution depend on sequencing sequences (Boeckx, 2017), which was proposed in the previous section as a pairing of cortical and cortico-subcortical circuitry, namely the dorsal pathway and the cortico-basal ganglia circuitry, human version of FOXP2 could strengthen the connections and fasten the processing speed of information which is under general learning process.

Enard (2002, 2009, 2016) has predicted that the human version of FOXP2 with two amino acid changes after splitting from common ancestor of chimpanzees could have been engaged in the vocal learning ability of human beings. However, findings from paleo-DNA have shown that this human version of FOXP2 had already been formed in the common ancestor of Human, Neanderthals (Green et al., 2010; Krause et al., 2007), and Denisovans (Reich et al., 2010). Thus, care should be taken in claiming whether this human version of FOXP2 is the key for language evolution, because whether Neanderthals and Denisovans possessed any form of language is still under dispute. Nonetheless, in a regulatory region of the FOXP2 gene, indeed there exists a difference between humans and Neanderthals and Denisovans, that is a binding site for the transcription factor POU3F2 (Maricic et al., 2013), suggesting that "some changes in FOXP2 expression, potentially relevant to spoken language, evolved after our split from Neanderthals" (Fitch, 2017). Furthermore, we find it interesting that the POU3F2 protein is linked to a single nucletide polimorfism, rs1906252, which is associated with information processing speed (Muhleisen et al., 2014). Incidentally, the POU3F2 is involved in the development of neocortex (McEvilly et al., 2002), and like many FOXP2 interactors, is also engaged in developmental and language delays like schizophrenia and autism, both of which exhibit language problems (Lin et al., 2011). Therefore, POU3F2 could be a critical factor for language development and evolution. This certainly needs further research in the future.

3.2 How speech is special?

How speech is special or specialized could be understood from two perspectives, that are speech perception and speech production. We believe that speech is special both ontogenetically and phylogenetically. On the one hand, existing evidence from developmental studies could be interpreted as a kind of support to the specialization of speech. On the other hand, comparative studies from psychology and neuroanatomy provide evidence in support of the proposal that sensorimotor integration in auditory-vocal modality presents a gradient among primates.

3.2.1 Speech perception

This section mainly discusses how speech perception is special. The question of whether speech sound is processed distinctively from other types of sounds is worth exploring. In the previous chapter, it was argued that vocal learning birds possess special brain areas (NCM and CMM) for auditory processing and auditory memory for conspecific songs. We should expect to find specific areas in the human brain for speech if the idea that speech is special is on the right track. In different circumstances, Chomsky has repetitively stated that humans share the same auditory system as nonhuman primates. He considers the fact that only humans are capable of selecting sounds out of noise as strong evidence for the critical role of the universal grammar in language acquisition. As an example, here is an excerpt from Berwick & Chomsky’s (2016) recent book (p 98),

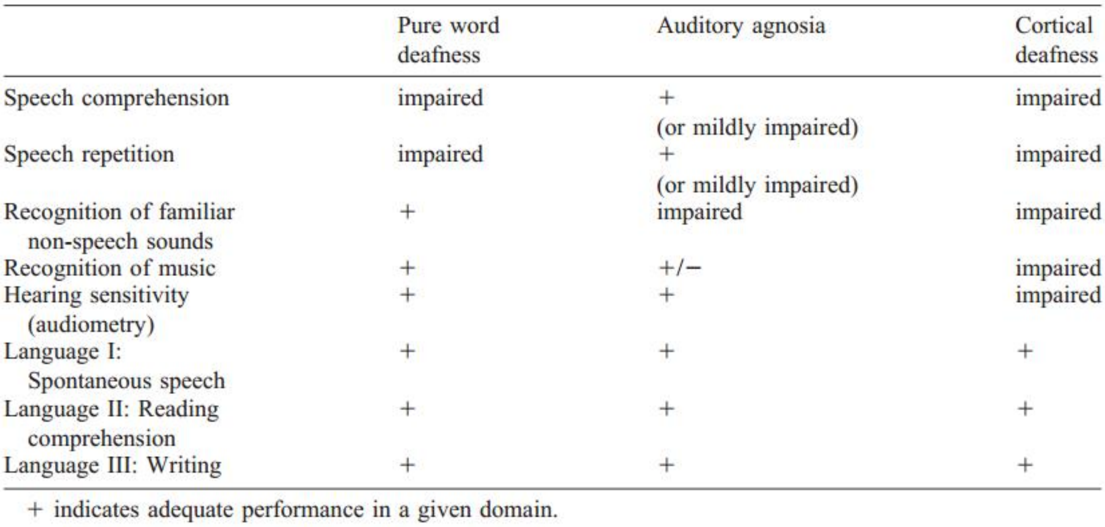
> "[...] which is evident from the first moment of birth. A newborn human infant instantly selects from the environment language related data, which is no trivial feat. An ape with approximately the same auditory system hears only

> noise. The human infant then proceeds on a systematic course of acquisition that is unique to humans [...]"

Chomsky’s view here is, however, in need of qualification. First, there is ample evidence proving that the auditory system is not the same in humans and nonhuman primates, at least quantitatively (see section 3for a detailed discussion). Second, the capacity to select the linguistic stimuli from the environment is certainly something related to vocal learning rather than to universal grammar --which is independent of any entertained view of UG. Now we will turn our attention to the reason for that. Newborns have no grasp of meaning, but they pay preferential attention to speech sounds. This will lead them to meaning and grammar later, but the initial selectivity they show for speech sounds seems to be just the same as the selective attention that all vocal learners show for the sounds emitted by their conspecifics. We showed in the previous chapter that in both vocal learning birds and humans, the processing of sounds is hierarchical, from the recognition of basic acoustic features to sound discrimination and categorization. We will present sources of evidence in the following sections supporting the idea that speech is special both ontogenetically and phylogenetically.

3.2.1.1 Speech perception is special---Poeppel

The specificity of speech perception differing from other types of sound processing has been proposed by Poeppel since 1990s. What the conspecific songs are for vocal learning birds is like what speech is for humans. Evidence from patients with auditory disorder suggests that the brain areas underlying the perception of speech sounds may be distinct from other sounds in general. The strong piece of evidence comes from pure word deafness (Poeppel, 2001), a form of auditory dysfunction connected with early hearing (e.g. frequency discrimination) and nonspeech auditory input including music but impaired spoken language comprehension. The patients' ability to speak, read and write are spared (table 3). The syndrome thus shows a neurophysiological double dissociation between speech and nonspeech comprehension, which suggests that with vocal learning circuitry in place, humans are instinct to absorb conspecific sounds.



	Pure word deafness	Auditory agnosia	Cortical deafness
Speech comprehension	impaired	+ (or mildly impaired)	impaired
Speech repetition	impaired	+ (or mildly impaired)	impaired
Recognition of familiar non-speech sounds	+	impaired	impaired
Recognition of music	+	+/−	impaired
Hearing sensitivity (audiometry)	+	+	impaired
Language I: Spontaneous speech	+	+	+
Language II: Reading comprehension	+	+	+
Language III: Writing	+	+	+

+ indicates adequate performance in a given domain.

Table 3: Auditory disorders following cortical and/or subcortical lesions (Poeppel, 2001).

With respect to the cortical areas for speech perception, it has been reported that spatially distinct subregions are responsible for musical instrument sounds, human speech, and acoustic-phonetic content, among which the left mid-STC shows selectivity for CV speech sounds as opposed to other natural sounds (Leaver &

Rauschecker, 2010). Another piece of evidence is more straightforward. Using sound quilts, which is a kind of stimuli preserving short timescale properties while disrupting long time scale ones, and functional magnetic resonance imaging, Overath et al. (2015) discovered that the superior temporal sulcus (STS) responds exclusively to the speech sounds. Manipulating sounds in various temporal scales, the authors rule out the possibility of amplitude modulation sensitivity or prosodic pitch variation sensitivity as factors interfering in the results. Overath et al. (2015) study suggests that speech is a kind of "humanized" sound that needs to be processed by special brain areas. Such idea is in parallel with the comparative studies on the specialized perception for conspecifics' sounds in vocal learning species, and auditory processing in nonhuman primates, which we will be discussed in details in the following subsection.

3.2.1.2 Specialized perception for conspecifics' sounds in vocal learning species evolutionarily--speech perception is special in humans

Exploring the validity of speech perception as special evolutionarily leads us to studies of nonhuman animals. In this section, we propose that in the evolution of language, the reason why vocalization successfully predominated other modalities as the modality of language production lies in the advanced auditory system of humans. Such proposal in turn favors the position that speech perception may be specialized in vocal learning species evolutionarily. Comparative data suggest that humans actually have more advanced auditory system than nonhuman primates. More broadly speaking, vocal learners are superior to non-vocal learners with respect to auditory ability.

Auditory learning differs from vocal learning in that the auditory learners never produce novel sounds, but are only capable of distinguish auditory inputs. There are plenty of species capable of auditory learning in nature. As our closest relatives, non-human primates are good auditory learners but poor vocal learners. Our belief that vocal learning could be the driving force of the predominance of vocalization over other modalities in vocal learning species including humans presupposes that there must be some cognitively and neurally qualitative or quantitative difference between vocal learners and non-vocal learners. In terms of perception, in vocal learning species, auditory inputs and feedback have been shown to be crucial in the process of learning tutor songs in birds and speech in humans. Therefore, the precision of the auditory perception seems indispensable, in the sense that at a minimum, it includes precise auditory discrimination, auditory detection and better auditory long-term memory. Taking these properties into account in human speech, they represent significantly enhanced auditory processing. Comparative studies suggest that non-human primates lag behind humans in auditory perception tasks, as for apes being worse than others and monkey being the worst. Non-human primates are less sensitive to lower frequency tones (Kojima, 1990) that are commonly present in human speech; they do poorly in auditory discrimination tasks (Kojima, 2003); and they have poorer auditory long-term memory compared with visual or sensorimotor memories (Fritz et al., 2005). Along the same line of avian species, the vocal non-learning birds are worse than vocal learning birds in all three aspects. Pigeons need more training time to acquire pitch discrimination than songbirds (Cynx, 1995). Some vocal learning species (zebra finches and budgerigars) achieve auditory discrimination ability even beyond that of humans (Lohr et al., 2006) (figure 21). Also, elephants that are identified as vocal learners are shown to be advanced in auditory perception (Heffner & Heffner, 1982).

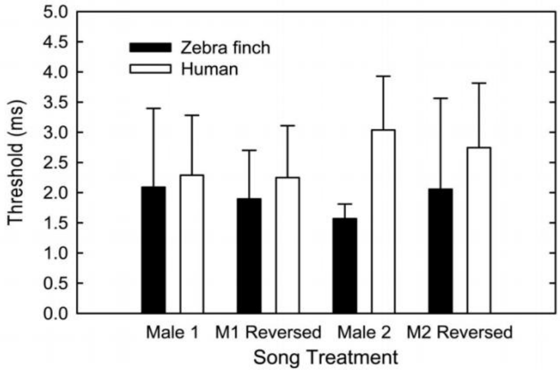


Figure 21: Comparative auditory ability between zebra finches and humans (Lohr et al., 2006).

Although with respect to the auditory neural pathway, there seems to be no difference between vocal learning birds and vocal non-learning ones (Jarvis, 2009), as we mentioned above, vocal learning birds (songbirds) have been identified with special nuclei for auditory memory, namely the caudal part of the medial nidopallium (NCM) and the caudal part of the medial mesopallium (CMM) (Bolhuis & Gahr, 2006) (figure 22). We stated in chapter 2 that in the human brain, the auditory processing structures are organized tonotopically and hierarchically (1.3.2.1), and speech processing is performed by a ventral stream dealing with phoneme and lexical recognition and lexical combinations, and a dorsal stream engaged in the sensorimotor transformation in language production (Hickok & Poeppel, 2004) (figure 23). The question that needs to be dealt with is the matter of auditory processing in nonhuman primates.

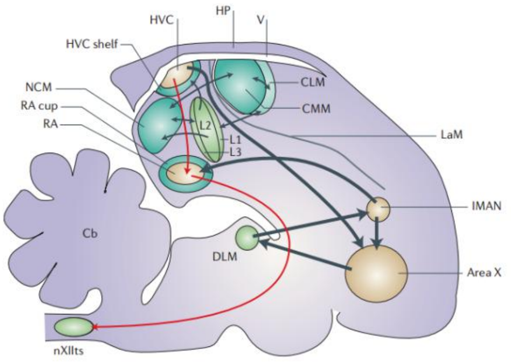


Figure 22: Brain areas for auditory processing in songbirds (Bolhuis & Gahr, 2006).

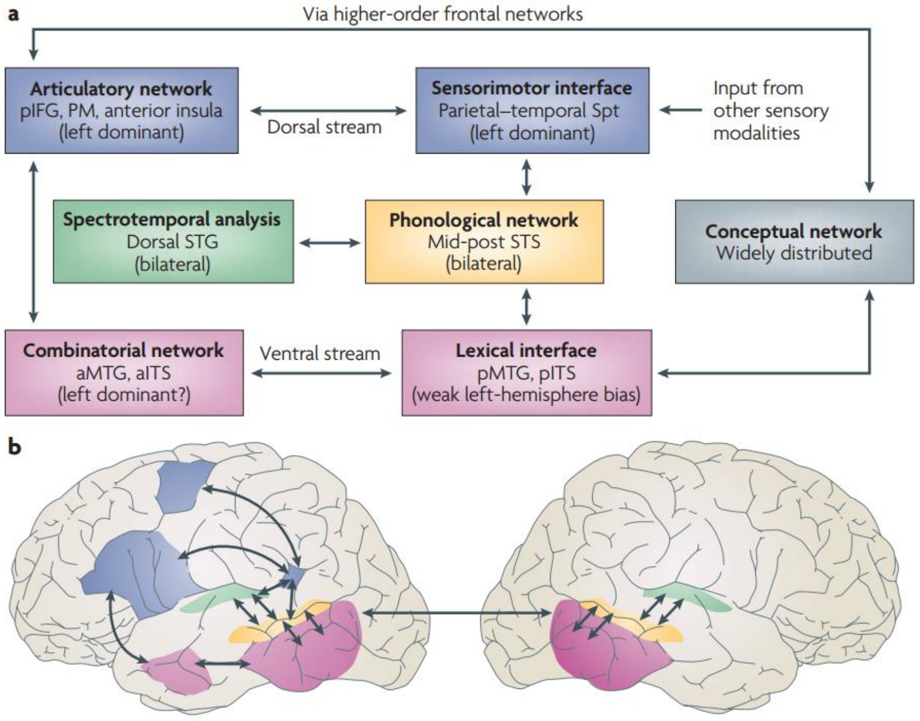


Figure 23: A comprehensive picture of dorsal/ventral pathway hypothesis of language processing (Hickok & Peoppel, 2004 et seq.).

Bornkessel-Schlesewsky et al. (2015) focused on the auditory ventral and dorsal pathways and proposed that the difference between humans and nonhuman primates is quantitative rather than qualitative. In case this holds true, it means that the more precise auditory perception ability of humans may be due to larger quantity or stronger connectivity of the neural connections between the auditory cortex and other brain areas. Studies have shown that the superior temporal cortex projects more massively and reciprocally to the premotor area and also more intensively to the neostriatum (Yeterian and Pandya, 1998; Rilling et al., 2008) than chimpanzees and monkeys (figure 24). This may potentially explain why auditory-vocal modality don't take over visual-manual one in non-human primates, leading to the absence of vocal learning ability in them. However, the neural basis provides humans with privileged skills for precise auditory perception, which is essential for vocal modality, and results in vocal modality bias with respect to other modalities in human speech, and in the selection of the auditory-vocal modality, i.e. speech, in evolution.

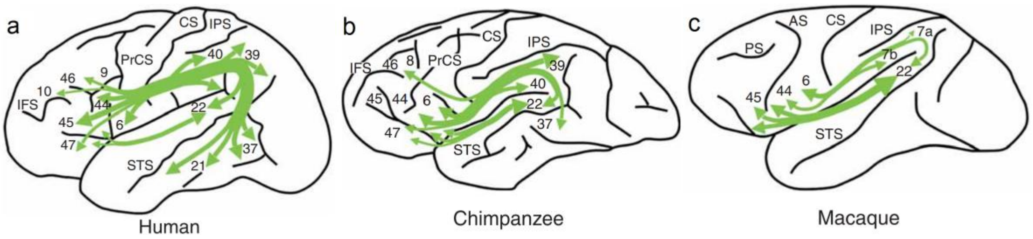


Figure 24: Humans have stronger and more enhanced connection from temporal cortex and other brain areas (Rilling et al., 2008).

3.2.1.3 Speech perception is special developmentally (ontogeny)

The perceptually selective ability of speech from noise starts from early infancy. Existing evidence has shown that the attentional preference for speech sounds by young infants indicates a speech biased initial setting for human beings

(Vouloumanos & Werker, 2007). This speech biased state observed in young infants, resembling vocal learning birds displaying preference to conspecific songs, suggests that vocal learning appears to take the lead in speech acquisition.

Speech which is inherited with properties, is a form of biological sound which is specifically produced by humans. It is also a sort of communication as well as linguistic signal. . Shultz & Vouloumanos (2010) attempted to investigate the level at which speech is preferred by three-month-old infants. The experimenters used nonnative speech, vocalizations of rhesus macaques (Macaca Mulatta), human involuntary non-communicative vocalizations, human non-speech communicative vocalizations, and environmental sound as stimuli. It was revealed that three-month-old infants listen longer to nonnative speech sounds than other kinds of stimuli. This shows that young infants attend selectively to speech, which suggests that humans may be endowed with a bias towards conspecific sounds, that is speech. However, children with language development disorders like autism spectrum disorder demonstrate atypical speech preference (Kuhl et al., 2005), which could be a main issue in play in their inability to acquire language normally. This will be closely analyzed in the next chapter.

It is also well known that newborns are evidently capable of discriminating distinct sounds of any language. However, as exposure to the ambient speech abounds, such capacity wanes gradually until children could only master discriminating contrastive sounds in their native language. Such phenomenon indicates that the auditory input of speech plays a crucial role in postnatal brain development---the neural development in hearing neurons, the thalamus and the auditory cortex, which are commonly accepted as auditory systems. The precise auditory discrimination ability has been attested in perinatal infants, which is proposed to be accomplished with a mature cochlea and brainstem, rather than the cortex since it is only mature in layer 1 (Moore, 2002). Kuhl (1992) have demonstrated that as early as 6 months, infants start to exhibit strong magnet effect for native language, but such prototypes in foreign language function as nonprototypes. Between 6 to 9 months, infants start to pay more attention to the high-probability syllables in their native language (Jusczyk & Luce, 1994). However, as they age, at 10 to 12 months, they fail to grasp the contrast between non-native phonemes. The phenotypes from 6 months to 12 months reflects "the maturation in thalamocortical afferents and an incipient participation of the deeper cortical layers (layer 4, 5 and 6) in the process of auditory perception" (Moore, 2002). The perception of masked and degraded auditory stimuli by children between 4 to 5 years and 11 to 12 years improves markedly, mirroring the maturation of superficial layers (layer 2 and 3) and corticocortical connections of the human auditory cortex (Moore, 2002).

Evidence supporting speech perception as ontogenetically special comes from the studies of the perceptual distinction ability of individuals who have had contact with a language during infancy, but to some reason adopt another language as native language later in life. Training in contrastive phonemes from Hindi or Zulu in native English-speaking adults in a short term revealed that those who had been exposed to Hindi or Zulu during childhood performed better in the perception task than those in the control group (Bowers et al., 2009). The same happens in the children adopted from India in America (Singh et al., 2011). Neuroimaging evidence is also revealing that infants from China who later are adopted in French-speaking Canadian families exhibit similar brain activation with the Chinese-French bilinguals in the discrimination task of Chinese lexical tones (Pierce et al., 2014). Finding an analogous effect of adoptees from Korea in the Netherland, and bearing the question whether the production will also be affected, Choi, Cutler and Broersma (2017) show that the benefits of the adoptees' perception of the language sounds could transfer to production. The data described above suggest that sounds are probably stored into the early memory, which could be unconsciously retrieved in later life, indicating that the auditory input in infancy could play a crucial role in language development, which in turn implies that the early exposure to the linguistic sounds could be one of the sources for the hard-wired brain.

3.2.2 Speech production

We have shown that speech perception is special both phylogenetically and ontogenetically. In the current section, we will discuss how speech production is special in phylogeny and ontogeny too.

Speech production has been studied in psycholinguistics with multiple levels of phonological computations and representations, and in the field of motor control orientation with phonetic realization or articulatory implementations. Combining two traditional theories from psycholinguistics and motor control, Hickok (2012) proposed a computational model for speech production. The architecture is separated into two levels, a higher level coding speech information at the syllabic level and a lower level of feedback control coding speech information at articulatory feature cluster level. The higher level involves a sensory-motor loop consisting of sensory targets in the auditory cortex and motor program in the BA44 and BA6. The area Spt (Sylvian parietal-temporal) coordinates the transform between the sensory and motor areas. The lower level involves a sensory-motor loop consisting of sensory targets primarily in the somatosensory cortex and motor program in the lower primary motor cortex. The cerebellar circuit mediates between the two. Hickok's model is greatly praised here as he has coped with the combination of psycholinguistic and motor studies. Nonetheless, in his early work, he only touched on cortical areas, rather than other neural correlates such as corticolaryngeal connection for speech production. We will discuss how BA44 and BA6 could affect the corticolaryngeal connection in evolution in the following subsection.

3.2.2.1 Speech production is special in evolution (phylogeny)

As a product of human biological evolution, speech production was proposed specialized owing to the peripheral factors such as descent of larynx. Afterwards, nonhuman animals were found to have the similar laryngeal descendance when they vocalize (Fitch & Reby, 2001), refuting the peripheral stand of the special status of speech production. Thus, it is inevitable to go deeply into the neural underpinning of speech production to explore how it is specialized from other nonhuman animals, at least nonhuman primates. The direct corticolaryngeal connection which is lacking in nonhuman primates, as was discussed in the previous chapter, was pinpointed as a key for speech production.

If this direct connection, postulated as “Kuypers/Jurgens laryngeal hypothesis” is the key innovation of humans among primates to produce voluntary sounds, it should be in some way connected with the cortical circuitry responsible for sensorimotor integration. Recently, Hickok (2016) has proposed that the Spt circuit has evolved in step with the direct corticolaryngeal control pathway, serving as the key innovation for human speech evolution. He provided evidence that there is more massive and stronger connectivity of the laryngeal motor cortex connecting with the inferior parietal and somatosensory regions in humans than in macaques (Kumar et al., 2016), and also made the interesting observation of approximated location between the inferior parietal target of the LMC and the Spt, and between Spt and the precentral sulcus with auditory-motor response properties. Nevertheless, in addition to Hikock’s hypothesis, we will add another important point to the evolution of brain areas, and that is, the subdivision of BA6 could have been a driving force in the enlargement of premotor cortex, where the direct cortical connection with larynx was generated over the evolution of speech production pathways.

As Chakraborty & Jarvis (2015) reviewed, the cortical complexity could be generated by gradual differentiation of a region into two or more areas. The expanded region could be developed into selective proportion carrying on new functions while the left parts maintain the original function. The ventral premotor cortex (PMv; BA6, boarding the BA44) of macaques has been shown to be responsive to visual, tactile and auditory stimuli and coactivate with the temporal lobe auditory cortex in the perception of species-specific calls (Stout & Chaminade, 2009). This suggests that in case of macaques in the evolution, the PMv was already responsive to the auditory-vocal modality together with the temporal lobe. In humans, the PMv is linked to the superior temporal gyrus and the orofacial motor cortex, forming a whole circuit for phonological articulation (Stout & Chaminade, 2009). It is only in humans that the PMv is divided into inferior and superior portions responsive to auditory and visual stimuli respectively (figure 25). The division mirrors the superior/inferior organization of hand and orofacial regions in adjacent primary motor cortex (BA4), where the laryngeal motor cortex is identified and corresponds to observed hand and mouth actions (Stout & Chaminade, 2009). This reflection suggests that BA4 and BA6 may have co-evolved in the evolution of human brains.

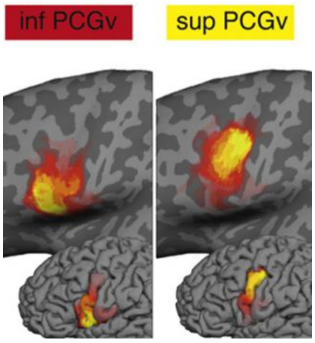


Figure 25: The specific subdivision of the ventral premotor cortex in humans.

This human-specific division of BA6 is implicated in language evolution. The division into inferior and superior regions of human BA6 makes each part specifically responsible for one modality. This in turn provides a platform for the sole auditory processing in humans. We hypothesize that this division of human BA6 is a driving force for the primary motor cortex expansion with the laryngeal motor cortex in place. This is in part consistent with Deacon's proposal (1989) described in the previous chapter, that more space is provided for the projection from the primary motor cortex to the motoneurons of larynx. As what Chakraborty & Jarvis (2015) argued, if this was the case, we would expect gene expression evidence to support our proposal. Furthermore, our hypothesis is that one of the human FOXP2 amino acid changes could have contributed to the posterior vocal learning pathway, i.e. the direct corticolaryngeal connection, hence there should be some link between FOXP2 and ventral premotor cortex in humans providing that our proposal held true.

Although no direct evidence has been found to show any sign of correlation between BA6 and FOXP2, a relative connection has been found in FOXP2 mutated KE family. The left premotor cortex is one of the overactivated brain areas in PET scanning of KE family members (Nudel & Newbury, 2013). Neuroimaging studies also show abnormalities of ventral premotor cortex in affected KE family members (Vargha-Khadem et al., 2005). This suggests that FOXP2 mutation is related to the normal activation and morphology of premotor cortex. Additionally, a recent paper has found that heterozygous mice (Foxp2-R552H, analogous to the R553H mutation discovered in KE family) display a posterior shift in the position and a more shallow peak in the distribution of the rudimentary laryngeal motor cortex (LMC) layer-5 neurons (Chabout et al., 2016). The data reviewed reveal that the premotor cortex and the laryngeal motor cortex mutually influence each other´s morphology and location, further suggesting that BA4 and BA6 may have co-evolved in the brain evolution. This area of inquiry needs further investigation in the future research.

3.2.3 Speech production is special in development (ontogeny)---babbling

Normally, speech production occurs later than speech perception in infants. As was shown in the previous chapter, when dealing with the analogy between the developmental stages of song learning in birds and language acquisition in humans, vocal learning appears to be an indispensable ability that enables the child to successfully acquire the ambient language(s) in the auditory-vocal modality. Confronted with the auditory signals as early as the auditory system starts to function, when still in uterus, the baby absorbs auditory input from the speech. When the baby is born, she is exposed to multisensory input with gradual maturation of her sensory system. Studies have shown that without experience, the neurons of such multisensory input cannot be generated properly (Tierney et al., 2009). This is what we discussed in the previous section about the maturation of the neural systems of speech perception.

In this section, we choose babbling, a crucial developmental stage of language acquisition, as our line of argument for the importance of vocal learning in language development. Babbling has been regarded either as a precursor of full-fledged language ability or as simply of vocal experimentation. It starts from birth, and develops in both acoustic and structural dimensions during the first year of life. In normal development, babbling is divided into the following stages (Oller & Eilers, 1988). In the first two months after birth, newborns start to make sounds like crying, coughing, sneezing, which do not involve vibrating vocal cords or any property of speech sounds. Together with these sounds, infants also produce a so-called quasi-vowel with some speech quality, which comes from the vibration of the larynx but not the rest part of the vocal folds. This is called the phonation stage. By two to three months of age, infants enter the gooing stage, where the seemingly precursive consonants are heard with the primitive movements of the articulators---the lips and the tongue. Moreover, the sounds begin to coordinate with eye contacts. During four to six months, which is called expansion stage, infants produce fully resonant vowel sounds and precursors of syllable that are termed “marginal babbling”. Laughter is featured in this stage. From seven months on, the canonical babbling stage starts. The recognized syllables with consonants and vowels which are the basic phonological blocks in language are heard at this stage. .

We believe that babbling is a crucial developmental stage for acquiring language, when infants experience uttering articulated sounds before producing recognizable words. Babbling is vocal production that involves learning, and the comparable stage can be found in vocal learning birds as subsong stage. The old-fashioned views about babbling might date back to Jakobson (1941), who proposed a discontinuity view concerning babbling and phonological acquisition. He posited that a child acquires sounds according to a set of systematic phonological rules with contrastive features, rather than imitating the ambient sounds, a view which is against the role of imitation in language acquisition. Lenneberg et al. (1965) also concurred with Jakobson's idea to some degree. He provided the illustration that deaf children also babble even if they themselves cannot hear their own babbling, and they start babbling at about the same age as normal hearing children do, but they stop babbling earlier than normal hearing children. Using such examples, Lenneberg tried to demonstrate that babbling does not need sound input and auditory feedback.

Nevertheless, the fact that the babbling of deaf children exhibits a variation that the quantity of babbling depends on the degree of their deafness (Oller, 2014), strongly suggests that the auditory input is critical in the early stage of babbling. Simulation experiments have also shown that auditory feedback plays a crucial role in the advancement of babbling (Warlaumont & Finnegan, 2016). Besides, the story of Genie (Curtiss, 1977) provides compelling evidence for this stand that without any form of auditory input, babbling as well as language will never emerge. Although as Lenneberg argued, deaf children also babble, they babble in a different way that no prominent syllabic structures occur. What they produce may be just some affective response with or without social stimuli. Further, the profoundly deaf children are reported to not babble at all (Oller, 2000), suggesting that auditory input is indispensable. Indeed, deaf children who learn sign language still babble, not in vocal modality but in motoric modality, with kind of prelinguistic sub-signs, which are high-skilled movements. It should be remembered that in case of absence of sign language exposure, deaf children will not be able to develop canonical babbling in motoric modality (Oller, 2000). Moreover, regardless of vocalization or movement, babbling exhibits language variations too (Oller, 2000). This supports the view that auditory (sign) input is important for babbling, and the view that babbling is most probably a precursor of language development.

The data suggest that the auditory input is necessary for babbling. We further propose that, if babbling is a required early stage of language acquisition, like the subsong stage in vocal learning birds, language will be initiated by auditory input and auditory-vocal integration. Neuroimaging evidence on deaf children seem to endorse our proposal. Surprisingly, functional imaging evidence has shown that the comprehension of the signs in deaf children activate auditory cortex (Finney et al., 2001; Nishimura et al., 1999), rather than visual cortex, although deaf children use visual-motor modality for language acquisition. What is more, congenital deaf children without cochlear implantation exhibit topographic tonotopy-based functional connectivity in the core auditory cortex including the language areas (Striem-Amit et al, 2016), further illustrating the importance of auditory involvement in language acquisition across modalities. The cases proving that the signs produced by deaf children need the activation of the auditory-vocal circuitry for normal hearing children indicate that auditory-vocal circuitry could have been the basis of language development, and it may be recycled for language acquisition in other modalities like signs.

3.2.4 Integration of speech perception and speech production: The role of vocal imitation

The perception and production of a complex behavior is integrated by imitation. It is well known that the repetitive practice of a skilled movement with the corresponding visual feedback is required for acquiring such movement. For example, if you want to learn how to make something like a pottery, you need to imitate the process step by step, and practice for several times with visual feedback, so that at last you can master such a high-skilled form of manufacture. Another example is learning how to play an instrument like the piano. Apart from imitation and visual feedback, auditory feedback also plays an important role, which is a cross-modal behavior. Speech is such a cross-modal product, in the sense that when imitating vocally, humans need to receive feedback in both auditory and visual modalities (McGurk effect). The traditional idea is that vocal imitation is derived and developed from general imitation ability. On the contrary, we propose that general imitation ability is an ability subserved from vocal imitation ability. Our proposal is consistent with Fitch (2000)´s speculation that vocal imitation might have preceded generalized mimesis in phylogeny. It should be emphasized Pay attention that true imitation usually involves both the means and the goals of the action, and by and large consists of multiple steps. Going to the neural level, the circuitry for vocal imitation could have been recycled for that of fine-grained general imitation.

3.2.4.1 Imitation in nonhuman primates

Imitation can be found at different degrees in nonhuman animals. Nonhuman primates are not vocal production learners. From the aspect of action imitation, nonhuman primates show more degraded imitation ability than humans. Upon the investigation of the action imitation in macaques and chimpanzees, it has been found that macaques are attentive to the transitive action (the action with results), whereas chimpanzees are attentive to the intransitive action (the action without results) like humans (Hecht et al., 2013). The transitivity of the action entails goal-directed intention in the macaque, while the intransitive action suggests that chimpanzees are capable of imitating specific details of an action, even if the action results in no consequences. The way the macaques combine discrete gestural elements into simple goal-directed actions are much like the coordination of the discreteness of the articulatory gestures to pronounce syllables in human language (Stout & Chaminade, 2009). This suggests that the ability to combine discreteness into a whole may have a common ancestor of both modalities. However, no evidence has ever emerged to show skilled action imitation in nonhuman primates, indicating that this high-skilled imitation ability is not present in primate lineage, but convergently evolved by humans and other animal imitators. If our hypothesis is on the right track, the vocal imitators should be good motor imitators.

3.2.4.2 Vocal imitation in humans

In the upcoming paragraphs, we attempt to contribute to the theoretical debate about the nature of imitation in light of vocal imitation, which systematically seems to be overlooked. Our position instead is that vocal imitation serves as a substrate for general imitation abilities.

It is uncontroversial that humans are the most flexible and skilled imitators in nature, in as far as they have been suggested to be called Homo Imitans species. We want to suggest that Homo loquens would be a more appropriate nickname for our species if a nickname should be chosen. In support of it, we will argue that our vocal imitation ability, which goes inherently with our loquens nature, is the basic imitation ability in our species and that other kinds of imitation rely heavily on it, as hinted by Fitch (2010). Furthermore, by putting vocal imitation as the irradiating center for general imitation, we advance a sort of a *tertium comparationis*, the Basic Vocal Imitation (BVI) proposal, between the two approaches to imitation that are most debated in psychology nowadays, i.e. the transformationalist one which is dominant, and the new associanist one. The former has been put forward by Meltzoff & Moore (1994) and argues that only our species, Homo Imitans, has an inborn Active Imitation Matching (AIM) mechanism that allows its members to align the representation of the self with the representation of a conspecific, when performing a given action, so that a topographic identity between both representations is obtained and real imitation can arise. In other words, the AIM is presented as a solution for the *correspondence problem*, meaning that an observer can see an action performed by another individual but cannot feel what the performer experiences in the action, whereas when copying another's performance with her movements, she cannot see herself in the performance.

The new associationist theory of imitation is the so-called Associative Sequential Learning (ASL) whose main proponent is Cecilia Heyes (Catmur et al., 2009). According to ASL, which is an associationist but not a behaviorist theory, imitation in humans does not constitute a radically new capacity. Apart from horizontal associations between visual and motor representations, the former is in charge of recognizing actions and the latter of performing them, ASL requires vertical associations between vision and motor representations to deal with the correspondence problem mentioned before. Vertical associations in these models are based on processing both contiguous and contingent events. By including the processing of contingent events, the model overcomes a shortcoming of the classical associative learning as the basis for imitation. The classical model, with contiguity as the only relation between events to build the associations, was too unrestricted as to explain how imitation could arise, since the associations would be too many for the observer to select among them, a point which has already been pointed out by Piaget (1952). By adding, as a necessary ingredient of imitation, contingent, predictive associations where the likelihood of a given second (imitative) event X is superior in case of a first instance of X performed by the observer, the necessary restrictiveness could be achieved for the relevant imitation event to happen.

Heyes (2015) has elegantly faced seven objections against her ASL model from the proponents of the AIM theory of imitation. Some of them, as Heyes shows, are not even supported empirically, despite claims to the contrary. This is so in the case of the first objection which says that newborns can imitate. This widespread belief must however be called into question. After Heyes' own questioning of the empirical validity of the experiments which have allegedly shown this, Oostenbrook et al. (2016) (a longitudinal study containing 106 infants) have completely undermined this claim on newborns' imitative abilities. This finding coheres with the default hypothesis that would follow from the BVI proposal, that is at birth there is no imitation in place since any vocal learner goes through a silent and next a practice stage (babbling) to match the model he gets from the tutor. Here BVI is on the same side as the ASL model.

The second objection stands that infants do not receive the right kind of experience —in a sort of a poverty of stimulus argument here. The AIM mechanism, instead, is designed to overcome this bad quality of the input. Its proponents seem to consider that even taking into account the role of predictive associations, the problem of infants not getting the right kind of experience persists. In particular, the proponents of AIM argue that the predictive ingredient in ASL model has an erosion effect that could preclude the formation of the vertical association, because if there were too many instances of a given event, X not followed by any instance of seeing X, this could impede the establishment of imitating X. For instance, too many instances of mouth opening happening in contexts with not seeing any instance of mouth opening would erode the establishment of an imitative performance of mouth opening. Heyes's response here is that, this criticism does not take into account the salience of a certain context to establish (or reinforce) a vertical association, an association able to prompt an imitative performance. For mouth opening, this more salient context would be observing an opening mouth in a face occupying the most part of the visual field of the observer.

Imitation of radically novel actions, imitation in animals, goal-directed imitation and improvement without visual feedback constitute the next objections Heyes discusses before the last one, which has a more methodological and metatheoretical flavor and consists of casting doubts on ASL because it would "steal the soul of imitation". We will not deal with all of them in detail here, but we want to briefly focus on those issues that have judged through the prism of vocal imitation. It might be illuminated and maybe dealt more satisfactorily with the BVI we proposed.

In doing so, we cannot help asking why vocal imitation is not even mentioned in this debate on the nature of imitation. Heyes concedes with her opponents from the AIM side that imitation is the key for humans. She concludes that we learn the gestures that makes us belong to a particular human group merely due to imitation. But do we not have to learn the words spoken in such a group primarily? And how do we learn the words? As even Hauser et al. (2002) were prone to accept, it seems that we learn them thanks to vocal imitation abilities that we share with other vocal learners, as shown extensively through the present dissertation. Whether vocal imitation is dismissed in the discussion of natural imitation in humans precisely because it is considered, as the words that it supports, conventional is a possibility which cannot be discarded. And if we were on the right track on this, this would show how misled one can be by extending the conventional nature of words to their mechanistic properties regarding sound processing. Instead, we consider such properties deeply rooted in the common abilities for sound processing of all vocal learners. Be that as it may, let consider the way exactly the BVI view can constitute a *tertium comparationis* between AIM and ASL views.

Vocal imitation puts human species in a continuity line with other vocal learners, for all of which vocal learning is genetically imprinted. This would then run against the view of humaniqueness consisting of a unique imitative capacity as affirmed by the AIM theorists. An independently and innately specialized ability, vocal imitation does not seem to require the vertical associations crossing across representations in different formats (visual and motor) in contrast to what is proposed in the ASL model. In addition, the reason why vocal imitation does not require vertical associations is that in vocal imitation, the sensory input comes out with (essentially) the same kind of representations as the motor output since both work on sound representations. This means that the sensorimotor integration fed by the hearing sense for the purpose of vocal production radically separates from the sensorimotor integration fed by sight for the purpose of movement production —with the special case of the signed modality in language, which will be discussed (section 3.3.1): only the latter requires the vertical associations proposed by the ASL model; or equivalently, for vocal imitation there is no correspondence problem as stated in the AIM theory. As was mentioned, vocal production in vocal learners is accompanied with total feedback for the utterer. The differences in the hearing are associated with the differences in the production and the association between the sensory experience and the motor one takes place in the same individual. The correspondence problem vanishes, therefore.

The fact that our species is endowed with multimodal imitation skills also follows from our vocal learning identity (for example, dancing, which we will discuss in section 3.3.2) . Moreover, that only vocal learners can follow the rhythm (sensory, auditory) with movements of their bodies (motor, non-vocal movements) but not vice versa, as there is no dancer that is not a vocal learner (Patel, 2006), is well compatible with the idea that vocal imitation is the basis of any flexible true imitation found in nature. In other words, other animals can imitate in a broad sense but true imitation and multimodality in imitation seem to be restricted to vocal learners, which stresses its key role.

Why any other species as imitative as ours are not found? Are our enhanced vocal learning abilities are entirely responsible for the huge difference in imitation abilities between us and the rest of species? The answer is no. Here we agree with Heyes, that technology (e.g. optical mirrors) and cultural enrichments as seen in rituals, drills and games have an important role. Yet, these cultural enrichments concur with speech in general and even outside these ritualized domains, speech provides us with the most specific tool to discuss the degree of (in)exactness reached by a given acquired equivalence correspondence, i.e. by an instance of imitation.

Therefore, vocal imitation, now as the necessary support for the words we use to discuss movements and locations, again makes a crucial contribution to the expansiveness of our skills in imitation.

Regarding the possibility of imitating “elementally novel actions”, the BVI view suggests that we can align with Heyes and be skeptical about the fact that in absence of verbal instructions, this feat has been really documented and can be done. Vocal learning is always based on vocalizations heard from conspecifics. Of course there is room for some novelty in the process of acquisition (the normal route of phonetic change), but this novelty cannot be radical and elemental. However, we disagree about the role of visual feedback which according to the AIM contenders, is not necessary to improve the imitative performance. Such an improvement is nonetheless interpreted as merely a rather more vigorous response from the ASL view. Be that as it may, it is indubitable that vocal learning as such does not require visual feedback, which in any case would be insufficient as the appropriate targets in the vocal tract remain hidden to the observer.

Finally, we align again with Heyes that mechanistic explanations do not “steal the soul of imitation” but on the contrary, make more fascinating how high cognitive abilities can be considered a recycled outcome of lower sensorimotor skills. Moreover, mechanistic causes are more in agreement with the nature of scientific explanation in general. As was asserted before, we think that descent with modification is more explanatory than other approaches, so that continuist proposals must be pursued preferentially. It is only as a last resort that we must abandon such proposals.

To conclude, we want to address the issue of the correspondence problem (AIM) or vertical associations (ASL) once more. Is it likely that the BVI proposal replace either options if all imitation was derived from BVI? It seems that empirical research is needed to decide on that. If empirical studies show that non vocal motor imitation and vocal imitation are independent in development, AIM or the vertical associations should apparently be maintained together with the non-transformational/non-vertical VI. Yet there would be a reason to cast doubt on the necessity of a parallel system like this one if there was a dependence which, we suggest, would put BVI at the basis, because it is simpler as it does not involve a matching of different representations. Empirically, this would correspond to an initial appearance of BVI followed by other kinds of non vocal motor imitation. Interestingly, this prediction is at odds with the most widespread belief in typical and atypical development research where vocal imitation skills are assumed to follow the development of gestural imitation. In this regard, research on the development of both kinds of imitation in normalcy and in autism should be highly informative. In this line, we will show in the next chapter that a close examination of the development of imitative gestures and vocalizations is in agreement with the BVI view and therefore, on a close relationship between gestural imitation and vocal imitation, with the latter as the trigger of the former.

Of note, the fact that vocal learning exists in all vocal learning species objects to the innatist stance of AIM and favors instead the ASL assumption with more consolidated empirical evidence. By doing this, the BVI puts vocal imitation as the core imitative learning ability in humans in the sense that the multimodal and multidimensional imitative skills would be promoted by vocal imitation.

Concerning the point that neural correlates for vocal imitation potentially engender those for general imitation, we will delineate it in the frame of multimodalities of language in the following section.

3.3 Multimodal evolution of language

In the previous sections, we put forward that auditory-vocal modality of speech is special in both perception and production in ontogeny and phylogeny. However, we do not believe that language has evolved from a monomodality which is auditory-vocal, but a multimodality with auditory-vocal one as the initiator simultaneously appended by other modalities. This idea makes speech not different from other communicative signals that nonhuman animals produce, that was argued in the previous chapter as multimodally evolved. In studies of language evolution, the question of whether gestures or vocalizations are the origin of language is still in dispute. However, no matter which was the initial modality, we have to explain how vocalizations occupied the position of dominant modality of language. It has been argued that vocal communication was selected because of cultural development. More specifically, vocalization was selected because of some advantage like using vocal modality did not disrupt manual work and sounds could also travel without light (Corballis, 2010). However, this intuition can barely explain the emergence of vocal learners in nature, who are not as much culturally developed as humans, but use auditory-vocal modality as the communicative channel. In this section, our hypothesis was that auditory-vocal modality is actually the very origin of language, not only for the sake of explanatory power, but also for the sake of parsimony in natural evolution. In the previous chapter, ample evidence was provided for the multimodal origin of animal communication signals dating back to vertebrates. In this section, our aim is to develop this multimodality point to the evolution of language and general cognition. As we have repeatedly stated, language has most probably evolved from existing abilities and co-organized them in a novel way, which Fitch (2017) call shared and derived components of language. The former could be found in other nonhuman animals, and the latter would be further developed in human lineage. In this section, by reviewing theories and relevant aspects of language evolution concerning modalities more than auditory-vocal, we further propose that the underlying neural basis responsible for auditory-vocal modality could have been recycled for other modalities and domains.

3.3.1 Gestural theory of language evolution

Although the gestural theory of language evolution has been well developed, we are not in agreement with it. We believe that language origin is multimodal, with the auditory-vocal modality as the basis, coexisting with other modalities. The primary gestural modality of language/communication evolution is rooted in the philosophical realm, and with the advent of psychological and neurological research, mounting evidence has emerged to be supportive, particularly the discovery of mirror neuron in macaques in the last century (Rizzolatti & Arbib, 1998). Mirror neurons are neurons that fire when the creature is performing an action as well as observing the same action performed by a conspecific (Hurford, 2002). It was originally discovered in the premotor cortex (F5) of macaques. The activation of the mirror neuron suggests a probable congruence between the execution of the action and imitation of the same action. Such discovery led to an assumption of gestural origin of language evolution, since F5 area seems to be homologous to Broca's area in the human brain (Rizzolatti & Arbib, 1998). However, although mirror neuron that is located in F5 area of the brain in nonhuman primates seemingly provided strong support for the gestural origin theory, the confirmation of the homologous existence of mirror neurons in human brain is still lacking (Hickok, 2009; 2014). This radically invalidates the logic behind the gestural theory of language origin.

On the other hand, the advocates of the gestural theory of language evolution attend to sign languages as another piece of evidence. Indeed, sign languages share properties with speech concerning reference, generativity, grammar, and prosody (Corballis, 2009), but this is far from adequate, because sign language is the only type of gesture that is symbolic, which is one of the significant features of human language, and other types of gestures in humans or nonhuman primates are devoid of it. Although apes like Kanzi or Nim Chimsky have been reported to be able to master some signs of American Sign Language or a large number of words in the channel of auditory mapping, the gestures the nonhuman primates produce are limited in number and variety, and that they have managed to learn words via auditory channel only shows that they are auditory learners, which we showed it in the previous chapter. Furthermore, it has never been found any form of gesture learning like vocal learning in nature. The parsimonious explanation in terms of evolution would be that language is set up on the basis of something existing in nature, but not a novel invention, since nature seldom works in this way (tinkering) (Jacob, 1977). Moreover, if gestural modality were the origin of language, why would vocal modality come to be predominant over gestures in human language, rather than language stay with the original gestural modality? It again deviates from the main biological principle of parsimony. Thus the gestural theory of language evolution posits a paradoxical discrepancy in the gestural theory of language evolution. We have stated our position that we consent with the idea that language origin is multimodal, with vocalization as the predominant modality, and co-occurrence of other modalities, like gestures. Broadly speaking, language is evolved with the advent of wide varieties of cognitive abilities re-organized in a unique way that only human beings possess. Vocal learning could have played a key role in this process, that is to say, it might have served as the basis of the advancement of general cognition.

The multimodality is also ubiquitous in modern communication. For instance, pointing gesture is under intense study because it requires joint attention, which is proposed by Tomasello as one of the distinguished features of human cognition. A recent study on pointing gesture synchronized with speech shows that the mismatch of the gestures and speech elicits enhanced activation of the left inferior frontal gyrus and bilateral posterior medial temporal gyrus (Peeters et al., 2017), which is consistent with the activation in McGurk effect that requires visual and auditory collaboration (Nath & Beauchamp, 2012). This in turn suggests that auditory input affects the language processing in a crucial way.

Beat gesture is another example which we deem is worth mentioning, though it is not widely studied. Beat gesture is synchronized with the rhythm of speech, especially the prosodic prominence. It was proposed that beat gestures serve as pragmatic significance. Studies have shown that beat gestures that have effects of increasing prominence could enhance the auditory processing of speech (Hubbard et al., 2009). The synchronization of the beat gesture and the rhythm of the speech needs to fit movement with time, the essence of which is similar to dancing. We presume that this beat gesture paired with speech shares the same origin of dancing as that in animals. As referred to in the previous section, birds dance with vocalization when they display courtship to females. The issue of dancing will be fully covered in the following subsection.

In summary, the gestural theory of language evolution cannot stand on its own. Language should have originated multimodally, with vocalization as the initial modality, and the neural substrate of auditory-vocal modality could have recycled by other modalities, like visual-manual gestures. We will review findings from other cognitive domains, supporting the idea of multimodal evolution of language as reorganization of different cognitive abilities, and further supporting our proposal that vocal learning ability may have provided the basis for high-level cognition.

3.3.2 Dancing

With the ability of synchronizing movement with time and the ability of imitation, dancing emerged in nature. Whether dancing is a unique human performance or is phylogenetically shared with other animals has been under a -long term debate. In human culture, dancing normally refers to the movements corresponding to the rhythm of some music. However, dancing can also occur intrapersonally, namely singing with dancing. Dancing might indeed be found among other nonhuman animals like the famous Snowball (parrot). Outside the human culture, if we define dancing simply as temporal imitating and synchronizing movements, this will also takes place in vocal learning birds, when they are trying to woo females with their wing movements synchronizing with vocalizations, the point which was discussed in the previous chapter.

Dancing is by no means an easy display, and requires entraining auditory and visual perception with motor production (Fitch, 2015). Scientists have long been curious about the evolution of such ingenious ability. Closer investigation of dancing of nonhuman animals reveals that the species that are capable of dancing are remarkably only the ones that are vocal learners. Owing to this, Patel (2006) put forward a "vocal learning hypothesis" that the vocal learning ability provides the vocal learners with more skilled ability of beat entrainment to account for such phenomenon. As we described in the previous chapter, vocal learning requires vocal imitation to learn conspecific sounds (song or speech), which is proposed by Laland et al. (2016) as key for dancing. In addition, Feenders et al. (2008)'s motor theory of vocal learning origin offers the neural evidence for this “vocal learning hypothesis”. Furthermore, the entrainment of the motor production to the beat resembles the beat gestures mentioned in the previous section, both of which require temporal synchronization of the rhythm between movement and sounds, which has been detected in vocal learning birds (ravens) (Pika & Bugnyar, 2011). All of this implies that simultaneous dancing and singing in birds, and concomitant beat gestures and speech may be dated back to a shared origin. This is in turn consistent with our proposal that vocal learning also plays an essential role in general cognition.

3.3.3 Co-evolution of tool making /use and language

Here we are returning to the motor aspect once again. Although we do not endorse the gestural theory of language evolution, we never deny the significance of probing the inquiry of language evolution from motor perspective. Tool making/use is a good example of exploring the role of motor aspect in the evolution of language as well as general cognition, and also the role of vocal learning serving as the basis of general cognition. Tool making/use requires multiple levels of highly integrated cognitive abilities, and has been treated as a golden standard to test overall intelligence by researchers. It has been explicitly expressed by Stout & Chaminade (2012) that the difference between language processing and tool making lies on primary sensory and the motor cortices, the intermediate processing is more overlapped. This implies that either one was the evolutionary basis for the other, or they co-evolved.

It has been hypothesized that there could be a co-evolution of tool making/use and language. The experiments on the Paleolithic stone tool making elicit the remodeling of the frontoparietal network (Hecht et al, 2014), which is overlapping with language processing network. The complexity of the tool making requires novel perceptual-motor specialization for visual and manual analysis, the development of executive ability for causal reasoning, and strategic planning (Hecht et al, 2014). This involves not only visual-motor hierarchy but also cognitive hierarchy which is abstract. The tool making activates frontoparietal, ventral premotor and the intraparietal sulcus, which are also activated in language processing. On the other hand, the Acheulean knapping, another kind of tool making technique of early humans, requires advanced cognitive integration, such as increased visuomotor coordination and hierarchical action organization, as well as working and planning memory. Acheulean world indicates that imitation and shared intentionality were already in place (Uomini & Meyer, 2013). However, the new functional areas for additional central visual field representations and increased sensitivity to the extraction of three-dimensional forms from motion are not found in monkeys. If tool making/use and language co-evolved, the implication would be that language is most probably the result of highly developed cognition, provided that the neural basis for both is overlapping. In other words, with the advancement of multiple cognitive abilities, language came into being with its unique organization in humans.

If our hypothesis that vocal learning serves as the basis for language and high-level cognition proved to be right, the non-vocal learners would not be able to master well enough tool making techniques. Macaques are not capable of using tools, even in very simple forms (Hecht et al, 2014). Chimpanzees are capable of making simple tools and using them, suggesting their ability of expressing intention, which gives rise to the expansion of anterior inferior frontal gyrus (BA45), but they cannot make complex tools with hierarchical multiple steps (Hecht et al, 2014). Among primates, it is only humans who are capable of complex tool use/making, and are the only vocal learning primates. Therefore, the complexity of tool making/use might be based on the complexity of vocalization in humans.

Logic suggests that vocal learners are expected to be superior than non-vocal learning peers in tool making/use, but this does not necessarily mean that any vocal learners should be better in using tools than any non-vocal learners. In other words, vocal learning birds should be more advanced than non-vocal learning ones, but certainly should not surpass nonhuman primates, since tool use/making also requires some cognitive abilities that may be present in nonhuman primates but absent in vocal learning birds. It has been discovered that many species of birds are able to use tools in the wild, on the occasions of reaching for food, as an example. The most famous tool makers in avian groups--crows, and a species of songbirds, have been found to make hook tools out of twigs and leaves in the wild (Hunt, 1996), and the complexity of the tools that they manufacture can rival or be superior to those of nonhuman primates (Hunt et al., 2006). Examples of other species of avian vocal learners include parrots that use small pebbles and date pits to capture calcium in shells (Megan et al., 2015). Furthermore, vocal learning mammals are good at using tools like Asian elephants (captive in naturalistic environments) that are capable of modifying branches, bottlenose dolphins that can use sponge as a foraging tool, and many other instances.

3.3.4 Action grammar

Another well-studied area in the motor domain is action grammar. We mentioned in the previous section that it is acknowledged in the linguistic theory that the core computation of language is recursive merge, which requires embedding sequences to sequences to form hierarchical structures. If our proposal that the sequences of sequences consist of one from vocal sequence learning and the other from existing dorsal pathway is on the right track, such sequences of sequences will never be detected in non-vocal learning species. We appreciate that more studies on nonhuman animals in cognitive domains other than language have shed light on the understanding of human language evolution. Action grammar is a recent reviving research topic for the purpose of exploring recursive merge in young infants and nonhuman animals in motor domain. Ontogenetically, three strategies have been observed in children development, which are pairing, pot and subassembly (figure 26) (Greenfield 1991). Studies have shown that nonhuman primates are capable of accomplishing the first two strategies, but fail in the third one (Conway & Christiansen, 2001), showing that the computational ability of nonhuman primates is limited to one dimension of sequences, rather than sequences of sequences. The subassembly strategy is proposed to be the hallmark of recursive computation shared with the computational ability in the domain of language. However, although the studies of such strategies on vocal learning birds are relatively rare, the use of subassembly strategy has been recorded as far back as the study of Herman et al. (1984) on bottlenose dolphins, where two captive dolphins were reported to be able to use subassembly strategy in treating surface pipe as a unit. Dolphins are vocal learning mammals, and it is conceivable that other vocal learners including birds and mammals could to some degree take advantage of subassembly strategy either in captivity or in the wild to solve problems. This prediction need further observation in the future research.

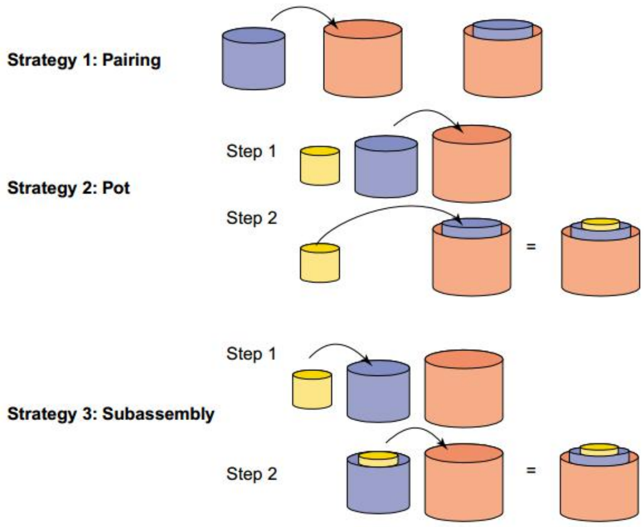


Figure 26: Three strategies. Paring requires to put a medial cup into a big cup; pot strategy requires a repetition of the pairing strategy, first to put the medial cup into the big cup, then to put the small cup into the medial cup that has already been in the big cup; subassembly strategy requires first to put the small cup into the medial one, then put the medial cup which contains the small one into the big one.

3.3.5 Summary

In this section, we focused on the multimodality of human language. In spite of disagreement with the gestural theory of language origin, we appreciated research from domains other than language to get enlightenments to better understand the evolution of language. We embraced the idea that language has evolved multimodally, in the sense that it has evolved cross-modally and across levels within cognition. We posited this to support the overall idea that language is a reorganization of cross-domain cognitive abilities.

4. Insights into autism from a vocal learning perspective

In this chapter, we will look at an atypical cognitive and communicative profile called Autism Spectrum Disorder (ASD) through the lenses of vocal learning. By doing so, we place ourselves at odds with the most recent Diagnostic and Statistical Manual (DSM), the DSM-5, according to which language if affected at all is a by-product of other causal factors. In previous versions of DSM, "gross deficits in language development" were instead in the definition of autism. In this dismissal of the causal role of language in ASD, the huge heterogeneity of the autistic population must have played a role. We contend, however, that even in autistic adults who purportedly have no linguistic deficit, subtle anomalies show up as sort of sequels of an abnormal development of language acquisition. Moreover, in terms of heterogeneity precisely, it cannot be overlooked that a 25-30% of ASD individuals are non- or minimally-verbal, a percentage that by far overcomes what is found in any other atypical/pathological condition. This portion of the ASD, however, is seriously under-researched which contrasts with the ever increasing research investments done in the verbal part of the spectrum.

A natural target to probe to what extent the neural mechanisms for vocal learning play a role in ASD would of course be the nonverbal part of the spectrum, in which speech (or sign), and language in the end, do not develop at all in a child who is neither deaf nor affected with peripheral problems as seen in apraxia of speech. Waiting for the empirical testing of this hypothesis, here we will argue that there must have been an atypical speech processing in early development across the whole ASD, which would amount to say that an atypical vocal learning in the human version is at the basis of ASD. We therefore will suggest that abnormalities in speech processing present in early (and later) stages of development (and later) play a causal role in ASD.

In order to fulfill this objective, we initially present a conceptual argument to replace language/speech at center stage in ASD, mainly at the crucial period of language acquisition. Next, we reinforce it with well-established evidence showing that processing sequential, fast and transient stimuli is more difficult or impaired in ASD. Afterwards, we deal with the inquiry of how speech production, perception and comprehension differ in ASD with a special focus in development (or language acquisition). Supportive evidence from genetics and neural correlates follows with the latter, having the potential to include repetitive behaviors as a manifestation of problems in the neural circuitry for speech. Finally, we argue that some light can be shed on abnormalities in imitation in ASD by considering vocal imitation as the basic imitative ability that feeds our general *imitans* nature.

4.1 ASD, speech, language and neurodevelopment

Autism spectrum disorder (ASD) consists of different conditions traditionally called Asperger's syndrome, pervasive developmental disorder not otherwise specified (PDD-NOS), autistic disorder and childhood disintegrative disorder. The "triad of impairment" (Wing & Gould, 1979) in ASD children is manifested in three domains: "reciprocal social interaction, abnormalities in communication, and patterns of non-functional restricted, repetitive and stereotyped behaviors" (APA, 2000). In the same vein, the DSM-5 lists, a compulsory component for an ASD, diagnose deficits in "social communication and social interaction" as well as "restricted, repetitive behaviors, interests or activities". At this point, even a naive observer could infer that atypical verbal communication, and then language must somehow be concurrent with problems in social interaction and communication. Language in the form of speech is the main vehicle and shaper of human social interaction and communication. If only because of that, it strikes us as surprising that language (or speech) is not even mentioned in the current ASD-5. Only by taking the so-called nonverbal communication as part of language, we could find a hidden, implicit reference to it. Equivalently, only in a broad sense will language indirectly be convoked into the deficits responsible for ASD. The DSM-5 does indeed mention a "poorly integrated nonverbal and verbal communication" but the whole weight is put in the nonverbal part as witnessed in references to abnormalities in body language (sic), eye contact and gestures that can be so severe to manifest themselves in total "lack of facial expressions and nonverbal communication".

How could one know, when integration of two parts is what is at stake, that only one part is affected? Since we have argued for the leading role of the spoken part in the multimodality of speech (see section 3.3), we should expect that speech itself, i.e. speech production and perception and comprehension in ASD presents us with difficulties. By definition, this is obtained in nonverbal and minimally verbal ASD, which seems forgotten in the ASD-5. Beyond the nonverbal end of the spectrum, there is however plenty of evidence that speech processing in the narrow sense, i.e. putting aside the non-oral ingredient, at the early developmental stages was (and is) compromised or at least atypical as we will see next.

Another striking point in the DSM-5 view of ASD is the dismissal of language acquisition in a disorder which is unanimously considered to be of neurodevelopment nature. Who can doubt about the impact of language acquisition in the development of the human brain? If "there is no real distinction between the development of the brain and the acquisition of language" (Balari & Lorenzo, 2015), as we agreed, how has a thorough assessment of linguistic development, i.e. speech processing from birth, been overlooked in the current clinical understanding of autism? Some influence on the DSM-5's position seems to come from Taylor et al. (2014), a huge twin study (3000 pairs) which found no correlation between language comprehension profiles and both phenotypes and genotypes. Actually the authors state in the abstract that their results "lend support to the forthcoming DSM-5 to ASC diagnostic criteria that will see language difficulties as separated from the core ASC communication symptoms". Interestingly, there is hope that the study may be flawed because it was based on mainly *written* tests, which do not present with the transient character that speech has, and that by definition the study was run on children well beyond the acquisition period. Moreover, if this study had some responsibility in the current DSM-5, it would be to some extent inconsistent with the own DSM-5 requirement that "symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies in later life)." (DSM-5:50)

Much more in agreement with the requirement that symptoms are present at the first stages of development is the following ASHA's (American Speech-Language-Hearing Association) recommended revision to the DSM-5. . It reads as follow:

> "Add a fifth diagnostic criterion for autism spectrum disorder: *Deficit in oral language"* [our emphasis].

A. Persistent deficits in comprehension and expression of language across contexts and modalities (e.g., spoken and manually coded), not accounted for by general developmental delays, and manifested as deficits in language form (phonology, morphology, syntax) and language content (semantics) ranging from limited language acquisition to total lack of comprehension and expression of language (as defined in section on language disorders).

Continue numbering the other criteria as B through E."

The recommendation, of course, follows a rationale. The elimination of the diagnostic criteria concerning spoken language would inevitably result in an inaccurate description of the fundamental nature of autism. With all language components including content (i.e., semantics), form (i.e., phonology, morphology, and syntax), and use (i.e., pragmatics, and social communication) in all modalities (e.g., oral and sign), language disorders are the hallmark of autism. It is obviously seen from the existing literature that spoken language disorders are a distinctive feature as well as an early key indicator of ASD. Furthermore, the generative system is absent even in verbal ASD children. Therefore, the inclusion of spoken language in the diagnostic criteria is needed.

In sum, we agree that language disorders are the hallmark of autism –that must be the reason why ASHA failed recommendation was presented as the A or first diagnostic criterion. As ASHA states, with language removed from the ASD "all children with ASD would also have to be diagnosed as having a language disorder because intrinsically, ASD encompasses language disorders" (Asha 2012: 11). Next, we will argue that the linguistic deficit is mainly that of speech processing.

4.2 Speech as fast processing

There is a reason why speech can be challenging in ASD. Speech is accompanied with sequential, fast and transient information processing, which is well known to be deficient in ASD (Noens et al. 2008). In this regard, we reference three different kinds of sources that converge to point out the difficulties, albeit at different degrees, that information processing presents to the ASD population when the stimuli fade immediately after being produced, so that a minimum processing speed would be required for processing to succeed.

First, for the nonverbal end of the ASD, it has been shown that augmentative alternative communication systems such as PECS (Picture Exchange Communication System) (Lerna et al., 2012) constitute a real means to obtain an elementary communication level which is useful to express needs but does not lead its users to declarative communication. For our purposes, it is enough to state that PECS is a non-transient, visual system which organizes information spatially.

Second, it has been shown that the Processing Speed Index (PSI) as measured by the WISC-III/IV (Wechsler Intelligence Scale for Children) correlates positively with communication abilities and negatively with communication deficits (Oliveras-Rentas et al. 2012; Rafael et al., 2013, among others). An apparent contradictory finding, that ASD children outperform, by being faster, those with typical development in Inspection Time (IT) tasks, has been elegantly resolved. PSI requires sensorimotor integration since it is assessed through two subtasks that require timed-motor responses in contrast to (IT) that lacks such motor component. It has to be stressed in this regard that PSI lower scores are found even in Asperger individuals who are undoubtedly those relying more on verbal rather than visuo-spatial processing.

Finally, there is a view on autism that in accordance with the magical world theory of the ASD points to an “impaired ability to detect probabilistic regularities over time” as an underlying key cause with the potential to unify the apparently unrelated symptoms (Sinha et al. 2014). This Predictive Impairment in Autism seems to have gained support recently (Cociu et al., 2017). Be that as it may, it is clear that it also posits problems in speech processing where probabilistic regularities over time are the rule.

4.3 Speech perception of ASD

This section directly addresses the speech problems that ASD children possess. As we described in the previous chapter, at the first stage, during the memorization (auditory) phase, auditory input plays a critical role in afterwords vocal learning, and further speech/language acquisition in humans. If at the beginning the auditory input is deprived to some degrees, the afterwards speech acquisition cannot be successful. Evidence from behavior, neuroanatomy, and neurophysiology has shown that ASD children indeed are impeded in auditory input (O'Connor, 2012).

With regards to behaviors, compared with typically developing children, ASD children commonly exhibit superior ability in pitch perception (Bonnel et al., 2003; Heaton et al., 2008). Surprisingly, those who display better pitch perception ability tend to have more language related problems (Heaton et al., 2008). Moreover, they show a hypersensitivity to loud sounds, but this sensitivity decreases with age. This indicates that it may not be the case that the more enhanced the auditory ability is, the better the vocal learning will be; however, there may be a threshold of the level of auditory ability in the auditory phase of vocal learning, within which the vocal learning species could manage to select conspecific sounds out of noise. The autistic kids present remarkably high-level ability of sound perception, which may prevent them from the innateness of bias towards conspecific sounds, leading to the failure for them to set up the auditory templates for future acquisition of the sounds. In another study, Kuhl et al. (2005) compared orientation of autistic children and typically developing controls. The results showed that children with autism turn their head greater to synthesized non-speech analogies than to motherese, whereas control group members exhibit equal head turns to both, suggesting that the normally manifested speech bias is largely degraded in children with autism. Furthermore, Groen et al. (2009) observed that children with autism performed significantly worse than controls in identifying two-syllable words embedded in non-speech noise, providing evidence of reduced ability of processing speech out of background noise.

With respect to neuroanatomy, Rojas et al. (2005) reported a lack of normal hemispheric asymmetries in the planum temporale in autistic children, which is a result of smaller grey matter volume of the planum temporale. Boddaert et al. (2004) detected that children with autism exhibit decreased grey matter concentration located in superior temporal sulcus (STS). In fifty ASD children, Gage et al. (2009) found the right superior temporal gyrus (STG) has larger volume compared with controls. Barnea-Goraly et al. (2004) observed that relative to typically developing children, the autistic children show reduced white matter integrity in superior temporal sulcus (STS) and medial temporal gyrus (MTG), two crucial brain regions for auditory processing. Wan et al. (2012) found a reversed pattern of asymmetry of arcuate fasciculus in nonverbal children with autism. Eyler et al. (2012) showed that infants and toddlers with autism display abnormally reduced left temporal cortex activity, and the toddlers exhibit reversed lateralized pattern in the anterior portion of the superior temporal gyrus, the pronounced brain region in response to sounds in typically developing children.

In terms of electrophysiology, Lai et al. (2011) found that in autistic children, there is less activation and activity spread in STG compared with children in controls in a passive listening task. Redcay & Courchesne (2008) noticed a pattern of lower left frontal-temporal activity but higher right frontal-temporal activity in ten sleeping autistic toddlers in passive listening to stories. This right lateralized pattern is observed in adolescents and adults with autism (Wang et al., 2006; Tesink et al., 2009;

Gomot et al., 2008), indicating that ASD individuals may take compensatory strategies. This requires further research in the future (O'Connor, 2012).

4.4 Speech production of ASD

Difficulties in speech production are the natural consequence of impeded speech perception. As was discussed in the previous chapter (3.2.3), babbling plays an important role in the development of speech production. Infants at risk of autism are temporally delayed in babbling, or do not babble at all in the population of nonverbal ones. This atypical initial stage of speech production will lead to future problems in producing speech. Echolalia, to repeat what others have said, commonly goes hand in hand with ASD children (Stiegler, 2015), but this character was considered as one type of repetitive behavior (Gernsbacher et al., 2016), that probably shares the same origin of deficits of vocal learning, a point which will be fully discussed in section 4.7. Moreover, children with ASD also manifest referential problems and pronoun reversal, the common example of which is their use of *you* for self-reference and *I* for an addressee (Evans & Demuth, 2011). Surprisingly, ASD children exhibit an atypical pattern of speech perception and production; in other words, that they are able to produce more speech than to comprehend speech. Provided that ASD children present speech perception problems as we described in the previous subsection, the consequence in the trajectory of speech development would be that they are equally bad at speech production, or even worse. It is thus conceivable that their relatively better performance in speech production could be obtained from a compensatory strategy which is not generated by speech system *per se*. This issue requires further study in the future.

Concerning the neural circuitry of speech production, we clarified in both chapter 2 and chapter 3 that the direct corticolaryngeal connection, the posterior vocal learning pathway in human version, is responsible for the production of learned sounds like speech and song. Although few studies have addressed the corticolaryngeal connection in ASD population, the point is implicated in genetic findings.. Wang et al. (2015) have shown that the forebrain part of the direct connection to brainstem of vocal learning birds has specialized regulation of axon guidance genes from the SLIT-ROBO molecular pathway. Coincidentally, Anitha et al. (2008) have shown that abnormalities of ROBO family may give rise to autism, which is a potential link between the posterior vocal learning pathway, i.e. the direct corticolaryngeal connection, and autism.

4.5 Sensorimotor deficits in ASD

The sensory and motor problems which arise from birth are prevalent in autism. The degree of severity of the sensory dysfunction in autistic people depends on the modalities tested (Hannant et al., 2016). Surprisingly, at low level of sensory perception, ASD children present equal or higher performance compared with TD children in visual domain (Mottron et al., 2006), whereas at higher levels, such as the global visual processing is atypical involving the dorsal pathway (Pellicano et al., 2004). In case of auditory modality, the results are similar. The dorsal pathways are known to be involved in sensorimotor integration across modalities. It has been demonstrated that the main sensory modalities (auditory, visual, touch, and oral) are not independent from each other, but they show a significant correlation (Kern et al., 2007). In addition, reduced ability of multisensory integration has also been found in ASD individuals (Stevenson et al., 2014). We know that visual and auditory inputs are integrated to serve the role of speech processing (e.g. McGurk effect). ASD children have been reported to show reduced McGurk effect (Bebko et al., 2014). As was described in the previous two chapters, that the dorsal pathway plays a crucial role in auditory-vocal integration and vocal production, the phenotypes apply perfectly to our hypothesis.

Evidence has also shown that ASD individuals have sensorimotor integration problems across modalities (Glazebrook et al., 2009), especially in multisensory coordination (Hannant et al., 2016). In motor learning, where sensory feedback is intrinsically connected with motor production, ASD children show difficulties too. However, with frequent practice, ASD children are able to improve the feedforward motor program (Hannant et al., 2016). Consenting with Hannant et al. (2016), we propose that the core problem of autism spectrum disorder lies in the general sensorimotor integration across modalities, but in the auditory-vocal as the basic one. It has already been proposed that individuals with ASD process information in a different way from normal developing ones. This in turn relates to fast processing that we explained in section 4.2. The widely used WISC for testing intelligence of autistic children reveals a discrepancy between cognitive abilities, in a way that the processing speed presents a lower score while the inspection time exhibits an equal performance compared with TD children (Wallace et al., 2009). The key point that accounts for this could be the fact that the inspection time task only involves perceptual ability, whereas the processing speed recruits sensorimotor integration.

Neuroanatomical evidence is also reported in this regard. Nonverbal children with autism show an atypical hemispheric asymmetry in the arcuate fasciculus (AF) (Catheine et al., 2012), the dorsal pathway for sensorimotor integration. Reduced volume of arcuate fasciculus is also found in high-functioning autistic adults (Moseley et al., 2016).

4.6 Genetic enlightenment from the study of ASD--CNTNAP2

In this subsection, we are going to provide genetic evidence for our proposal that malfunctioning of vocal learning ability may be the underlying reason for ASD. By focusing on findings concerning cntnap2, which is both linked to vocal learning and autism in humans and animal models, and with the argument we made on the vocal imitation and high-skilled imitation, we assume that the deficits in vocal learning coming from cntnap2 give rise to the abnormality of ASD.

Contactin associated protein-like 2 (CNTNAP2) is a gene that is closely related to vocal learning. It is "located on chromosome 7q36, and it codes for a neurexin superfamily member, whose functions in the nervous system are relevant to cell-cell interactions and ion channel expression (Poliak et al., 2003)." It is highly expressed in the frontal and anterior regions of the cortex that correspond to cortico-striato-thalamic circuitry (Alarcon et al., 2008) associated with the anterior vocal learning pathway. Moreover, evidence from birds is also revealing. In zebra finches, Panaitof et al. (2010) reported that punctuated expression of Cntnap2 was observed in key song control nuclei in males. Condro & White (2014) also found that Cntnap2 protein is enriched in song control regions in adult males, particularly robust nucleus of the arcopallium (RA). In addition, it is reported that CNTNAP2 is a downstream regulatory target of FOXP2 in humans (Vernes et al., 2008), and recently, Adam et al. (2017) have shown that CNTNAP2 is also a direct FoxP2 target in songbirds.

Coincidentally, CNTNAP2 is also linked with autism. Converging evidence has been provided on potentiality of the risk to ASD or ASD-related endophenotypes from common and rare variation in *CNTNAP2* (Peñagarikano & Geschwind, 2012). Alarcon et al. (2008) identified a variant in *CNTNAP2* (rs2710102) which was pronouncedly in linkage with the age at first word among ASD affected males. The same variant has also been reported to be associated with nonword repetition in language impairment (Peter et al., 2010) and dyslexia (Vernes et al., 2008). Interestingly, the preliminary results of behavioral experiment on nonword repetition in ASD children also suggest that this population performs worse than the control group (unpublished data). Furthermore, mutant mice model has become a practical tool to investigate the neurological and behavioral characteristics associated with ASD. The evidence from mutant mice model is also compelling. Cntnap2 mutant mice exhibit deficits in all deficient behavioral domains of ASD populations, that are "reduced vocal communication, repetitive and restrictive behaviors, and abnormal social interactions (Peñagarikano & Geschwind, 2012)". Moreover, the Cntnap2 KO mice show impairments in spectrotemporal auditory processing, which is consistent with the idea of complex sensory abnormality related to ASD individuals (Truong et al., 2015). Riva et al. (2017) investigate the role of rapid auditory processing (RAP), which is an early predictor of language development (Piazza et al., 2016), in ASD individuals. The results show that RAP functions as a mediator between the variant rs 2710102 and early expressive vocabulary, associating CNTNAP2 with auditory processing.

In conclusion, genetic findings additionally support our proposal that the malfunctioning of vocal learning ability could have been the underlying reason for deficits found in ASD populations. Cntnap2 causes the two streams of evidence--the role of cntnap2 in vocal learning and the role of cntnap2 in autism—to converge, which implies that at genetic level, vocal learning and autism are closely related.

4.7 Repetitive behavior and the atypical development of the basal ganglia circuit

We already proposed that vocal learning serves as the basis for cognition. In case this proposal was true, we would expect that the phenotypes of cognitive impaired disorders to be derived from the deficits of vocal learning. Restricted and repetitive behaviors is one of the distinguished hallmarks of autism. It is superficially distinct from the other two because it does not concern language. However, in this subsection, we will argue that the repetitive behaviors could also be a phenotype derived from deviant development of vocal learning. It has been proposed that this atypicality in terms of motor could have originated from the abnormality of the basal ganglia (Calderoni et al., 2014). We have stated in the previous chapter that the striatum is one of the vital areas involved in the anterior vocal learning pathway. The anomalous development of the striatum will induce aberrant development of vocal learning. We propose that the phenotypes of repetitive behaviors and atypical development of vocal learning in ASD individuals could have generated from the same origin, namely the atypical development of the striatum. Studies have shown that there seems to be a relationship between the restricted and repetitive behaviors and the volume of the basal ganglia. Enlarged volumes in right caudate and putamen have been observed by Hollander et al. (2005) in autistic adults. Langen et al. (2009) have reported that the striatal development with the volume of caudate nucleus increasing with age is deviated from typically developing control group subjects whose volume of the caudate nucleus decreases with age.

4.8 Imitation in ASD children

In chapter 3, we presented the BVI view which posits that vocal imitation is the basic one, as a *tertium comparationis* between the maybe two main theoretical approaches to imitation, that of imitation being sort of an innate module in humans, the AIM (Meltzoff & Moore, 1994), and imitation not being a radical novelty in humans but the development of preexistent abilities in other animals as defended by the ASL (Catmur et al., 2009). As was mentioned, since vocal imitation is present in all vocal learners, it opposes the foundational AIM assumption and confers the other side of the debate, the ASL, with a much solider empirical basis. As such, vocal imitation is placed as the core imitative ability in humans in the sense that the multimodal and multidimensional imitative skills would be prompted by a basic vocal imitation in the species. These are phylogenetic considerations though.

Regarding ontogeny, the BVI predicts a general impairment in imitation among the autistic population. The rationale behind this claim is that viewing the abnormal linguistic development found in ASD, as reviewed above, as the outcome of an abnormal speech processing system at early (and later) stages of the neurodevelopment would be equivalent –according to the position defended here– to targeting the human version of a more general vocal learning mechanism as a main factor for the linguistic abnormalities. Since vocal learning is inherently linked with vocal imitation, according to the BVI hypothesis, a general deficit in imitation in ASD is expected, which seemingly turns out to be right.

Vivanti & Hamilton (2014) in *Imitation in Autism Spectrum Disorders*, which is a review of the topic that constitutes the chapter 12 of *Handbook of Autism and Pervasive Developmental Disorders* (Volkmar et al. 2014), conclude the section entitled 'Imitation in ASD: Findings' with the following paragraph:

"In summary, current available evidence suggests that individuals with ASD, as a group, imitate others less frequently and less accurately from infancy, at least when compared to typically developing peers. Despite gains over time in imitative abilities, they continue to show impairments throughout the lifespan. These impairments are more obvious in tasks that measure true imitation, that is, copying the demonstrator's actions and goals without relying on knowledge about the outcomes of the action or the function/use of materials involved in the demonstration. In contrast, individuals with ASD seem to imitate better when tasks involve objects, when they are familiar with the materials involved in the task, when they understand the demonstrator's goals, and when they are interested in the outcome of the action. Moreover, imitation of single actions seems to be easier in this population than imitation of sequences of actions. Differences in imitative behavior appear to be associated to differences in social, communicative, as well as motor skills in this population; however, the nature of these associations is still not clear. Imitative difficulties are unlikely to play a causal role in autism, given that not all individuals in the spectrum show an imitation impairment and at-risk siblings who do not develop autism show a comparable deficit in imitation in infancy (G. S. Young et al., 2011)" Vivanti & Hamilton (2014:285).

Could it be the case that ASD individuals have a general imitation ability deficit? This is what Vivanti and Hamilton (2014) suggest without even mentioning vocal imitation. In accordance with the general neglect of vocal imitation in theories of imitation, we already mentioned (in section 3.2.4.2), the phrase *vocal imitation* does not appear in the paper despite the fact that vocal imitation fails in the 25-30% of those that are nonverbal in the spectrum. At this end of the spectrum, on the other hand, children and adolescents alike are almost or completely unable to imitate non vocally as well (research in progress by the Grammar and Cognition Lab). This by itself strongly suggests a link between vocal and non-vocal imitation.

Vocal imitation, on the other hand, conforms perfectly well to those patterns that come out as the most affected in the quotation above. First, vocal imitation is the quintessence of true imitation. The sequence of sounds that constitute the vehicle for words must be imitated by the learner with exactness, otherwise lexical acquisition would not be possible. What the rationale is for the sequence of segments that enter a word is nothing less than convention, which in turn entails imitation. That is what Sausurean arbitrariness captures. True imitation, to put it differently, is necessary to build a meaning, and a word. Second, vocal imitation is inherently a sequential action, which is most difficult to imitate than a single one in ASD. Third, vocal imitation takes place in the absence of an object. In this regard, it is like gestural imitation which, in contrast to vocal imitation, is mentioned and presented as more impaired than a kind of imitation that goes with or is performed on an object. Finally, vocal imitation has no paragon as far as its social function. It is again surprising that the authors state that humans tend to engage in imitation to establish bonds in social groups, but in this context, speech and the vocal imitation that supports and goes with it are not even called for.

To stress our point, we will briefly show that theoretical accounts of both autism and imitation have shortcomings to explain the findings about imitation in ASD.

A strongly supported finding, according to the comprehensive review of the literature, in the chapter in question is that ASD individuals are much better at imitating actions with a goal (emulation) than imitating for the sake of imitation (true imitation). This imitative performance does not fit the prediction of one of the so-called Weak Central Coherence theory of autism (Happé & Frith, 2006). The theory have predicted the opposite, because it denotes that ASD would be the result of the incapacity to build a complete account of what is perceived. Autistic individuals would be lost in the details for which they can even have an enhanced intake. The point that goal-directed imitation in ASD is much more preserved than pure imitation is clearly in contradiction with the theory of Weak Central Coherence. This in turn is also inconsistent with the closely related proposal (also Happé & Frith, 2006) which appeals to poor visual encoding as the source of the imitation deficit in ASD. In this case, the just mentioned finding is inconsistent with the idea that in ASD there is a bias toward the subcomponents of the visual stimulus which impedes global processing.

Other “theories” that have been proposed to explain the imitation failure in ASD presented by Vivanti and Hamilton are of the kind we disfavor because of their top-down nature. They are the “failed direct self-other mapping” which is to a great extent the same as the Mirror Neuron Theory applied to autism. It is a theory that apart from other shortcomings it presents (Hickok, 2014), faces the same counterexample as the weak central coherence when applied to imitation in ASD. In fact, it predicts that with a broken mirror neuron system, which is a main ingredient of autism, emulation (the copy of the goal) should be impaired because understanding of goals is compromised by such a broken system.

Another top-down attempt to explain why imitation is difficult in ASD points to an “abnormal social top-down control”. In this case, as Vivanti and Hamilton recognize, the limitation is that in this account the basic (bottom) imitation mechanisms would be intact. The finding that does not cohere with this account is now that accuracy is bad in ASD imitation. As was asserted before, this instead is a perfect fit for the BVI view.

Is the sensorimotor deficit account best positioned in explanatory power? The answer is seemingly positive. The authors list the following facts in support of the above answer:

> “(1) evidence that dyspraxia is common in ASD (Mosconi, Takarae & Sweeney, 2011; Rapin, 1996), (2) evidence that imitation accuracy in this population decreases as the motor demand increases (e.g., in sequential versus single action imitation tasks), (3) evidence of associations between levels of motor and imitation abilities in this population, and (4) evidence of an

association between imitation performance and abnormal visual-motor integration specific to ASD (Izawa et al., 2012)" (Vivanti & Hamilton, 2014, p 290).

In addition to more explanatory strength, the abovementioned account is the only one in line with ours, which is by definition a sensorimotor account as well, with a difference that it is based not on the visual system but on the auditory one for the reasons deployed in this dissertation (section 3.2.4.2).

Finally, it is worth mentioning that the neural model of imitation proposed by Vivanti and Hamilton (2014) greatly overlaps with speech/language areas in the brain. It comprises STS, IFG, MTG, IPL and mPFC, where all areas except the last are uncontroversially central in speech (see section 3.1.1 and 3.1.2 in chapter 3 of the dissertation).

In sum, we have shown that our predicted imitation impairment exists in ASD. Moreover, it has been demonstrated that current accounts of both ASD and imitation separately fall short in terms of their explanatory strength. The hypothesized central role of vocal imitation is then reinforced, but further research should clarify how general imitation has been built on it and how both of them work and interact in behavior and brain terms.

4.9 Summary

In this chapter, we presented evidence to support our proposal within the scope of an atypically cognitive developing population, that is, individuals with autism spectrum disorder (ASD). We firstly laid a conceptual foundation of speech being central in ASD. Then we showed that ASD individuals have difficulty in fast sequential processing, which is a cross-domain requirement for speech processing. We next provided evidence of anomalous speech perception and production in ASD with genetic, neuroanatomical and neurophysiological supports. In addition, we suggested that the sensorimotor problems found in ASD could be derived from sensorimotor problems in auditory-vocal modality. Our position furthermore explained the phenotype of repetitive behavior in relation with speech processing problem which both could originate from the atypical development of the basal ganglia. Reviewing Vivanti & Hamilton (2014), we finally argued that our proposal claiming vocal imitative ability as the basis of other imitative skill is more explanatorily adequate for the imitation deficits seen in ASD.

5 Conclusion

The present dissertation discusses the role of vocal learning in language development and evolution. We present ample evidence supporting our main hypothesis that vocal learning ability most probably lays the foundational basis for both language development (ontogeny) and evolution (phylogeny), and then is also contributing to high-level cognition. To our knowledge, our work pioneers the exploration of vocal learning as an essential contributor to language and cognition. Although compelling evidence directly addressing the question is still missing by integrating existing evidence on different aspects of vocal learning from the perspectives of development and evolution in both humans and nonhuman animals, this auditory-vocal capacity (recycled in the signed modality) manifests itself as indispensable for language, communication and cognition. Weare confident that more pinpointing evidence in the same vein we have presented here will emerge in future studies.

In chapter 2, we review current issues of vocal learning on animals including humans from diverse dimensions and across multiple levels. Analogies between vocal learning birds and humans covering developmental stages, neural pathways, genetic underpinnings, and evolutionary trajectory have been presented as shedding considerable light on the nature of vocal learning (Jarvis, 2007, 2009; Fitch & Jarvis, 2013; Pfenning et al., 2014). Additionally, we come up with several hypotheses that are in need of more empirical research. The first hypothesis concerns the evolutionary order of the anterior vocal learning pathway and the posterior vocal learning pathway. Although the evolutionary route of the pathways has been properly proposed (Feenders et al., 2008), no research has been done yet on the evolutionary sequence of the two vocal learning pathways. With comparative evidence, we hypothesize that the posterior pathway could have evolved before and served as a prerequisite of the anterior pathway. This would in turn strongly suggest that in language evolution the direct corticolaryngeal connection found only in humans among primates should have evolved before the cortico-basal ganglia-thalamo-cortical connection.

In the process of reviewing, we found that few studies have been done on the role of the hippocampus in vocal learning. In this regard, it is accepted that the hippocampus is involved in memory (Squire, 1992; Tulving & Markowitsch, 1998), which suggests that it must play some role in vocal learning. We propose that the hippocampus could be involved in the memorization stage of vocal learning and the extended song learning phase in open-ended vocal learning birds. We identify evidence on the positive correlation between the hippocampal volume and the continuum of open-endedness of vocal learning birds (Brenowitz & Beecher, 2005; Devoogd et al. 1993). To have a better knowledge of the involvement of the hippocampus in vocal learning may be consequential for the study of memory systems and in particular for the role the hippocampus plays in human language.

The human version of FOXP2 containing two amino acid changes since splitting from the common ancestor with chimpanzees has been proposed to play an important role in the emergence of vocal learning in human lineage (Enard, 2001 et seq.). We hypothesize that these two mutations play separate roles in the formation of the two vocal learning pathways: being one (T303N) for the anterior pathway and the other (N325S) for the posterior one. Only one of the mutations (T302N) in mice model (analogous to T303N in humans) leading to the morphological change of the striatum (Enard et al., 2009) is a revealing evidence supporting our proposal. More studies need to be done so as to tell whether our proposal is true or not. This research line will be beneficial to the study of genetic basis for vocal learning, in particularly in humans, which will in turn enlighten the evolution of language.

We further hypothesize in chapter 2 that foxp2 plays a more generic role across domains in language and cognition. By reviewing comparative evidence in drosophila, mice, birdsong learning, speech disorders, and multimodality in other cognitive domains, we draw a conclusion that foxp2 cannot only be a language gene or the language gene, but a gene that plays a critical role in multiple domains in general cognition.

We finally propose in chapter 2 that breathing could have played a vital role in the evolution of multimodal communication. The specific respiratory rhythm in humans has been described by MacLarnon & Hewitt (2004) with fine grain-precision. In line with it, with the findings that the periaqueductal grey (PAG) regarded as an area responsible for respiration (Subramanian and colleague's work), as well as an important primitive sound production area in the brain (Jurgens and colleagues' work), we hypothesize that the origin of multimodal communication signals may come from breathing. This hypothesis brings the multimodality to the very basic physiology of human, which could be too simple to explain such phenomenon. However, this may turn out to be how nature works.

In chapter 3, we explore the role of vocal learning in language, and furthermore general cognition. Once meaning is subtracted, vocal learning boils down to speech in the study of language. Even recursion, the specific computational operation of language, falls in with the workings of an enhanced vocal learning ability. Linguists have pinpointed recursive merge, the embedding of a phrase within a phrase, in slightly different views. In Boeckx's view, recursion could be generated by pairing two sequences together (Boeckx, 2016). We propose that the neural basis of one of the two sequences came from the basal ganglia circuit of sequence learning already well-formed for vocal learning, and the other came from the dorsal pathway of sensorimotor integration in auditory-vocal modality. FOXP2 and POU3F2 could have contributed to such pairing.

In line with the current view that speech is special (e.g. Poeppel, 2001), we present evidence arguing that speech is special both in perception and production in both development and evolution. The superior temporal sulcus (STS) has been identified as specifically responding to speech sound. Comparative evidence has shown that humans possess stronger and more massive connection between the temporal cortex to other brain areas (Rillings et al., 2008), which may be a key to more advanced auditory ability in humans among primates, leading to a predominant role of auditory-vocal modality in human language evolution. Moreover, newborns exhibit a speech-bias towards environmental sounds (Vouloumanos & Werker, 2007). The cases of adoptees who had contact with a language during infancy but speak another language as native when they grow up (Choi, Cutler & Broersma, 2017) is also a strong piece of evidence for the leading role of auditory input as imprinting effect.

In addition, we hypothesize that the specific subdivision of the ventral premotor cortex (BA6) in humans among primates may be an evolutionary driving force for the enlargement of primary motor cortex (BA4), where the laryngeal motor cortex that monosynaptically connects the larynx for voluntary vocal production. Ontogenetically, we take babbling as an example to argue that as an embryonic form of vocal production, babbling is a crucial developmental stage for speech acquisition and consequent communicative and cognitive development. Going to the sensorimotor integration required to learn to speak, vocal imitation is inevitable to be discussed. We hypothesize that vocal imitation could be the basis for the high-skilled imitation seen in humans in other domains on the basis of a comprehensive review of Heyes (2016), further proposing the idea that we humans are *Homo loquens* primarily and *Imitans* secondarily, thanks to our vocal learning abilities.

We also discuss the point that, speech as communicative signals in humans also multimodally evolved. The multimodality presented in speech evolution is not confined to the external modalities like auditory-vocal or visual-motor, it also contains the interaction between multimodalities from different cognitive domains. This leads us to be expectant on the research trend of exploring the evolution of language from a perspective focused on motor abilities potentially related to high cognition. In this regard, ideas of co-evolution of tool use/making (Stout & Chaminade, 2012) and language as those presented in the proposal of action grammar (Greenfield 1991), etc. cannot be ignored.

Chapter 4 is sort of a preliminary exploration of autism, a neurodevelopment disorder that also affects learning, from the perspective of vocal learning, a kind of learning that seems to be basic for human typical cognition and communication. By doing so, we implement a maybe stronger version of what was the rule in previous visions of ASD and corresponding diagnosis criteria, namely that in ASD deficits in language are core to the disorder. Core behavioral features that go with vocal learning seem to be absent or abnormal in ASD, to variable degrees depending on the severity of the autistic profile: from lack of preferential attention to speech to abnormal or nonexistent babbling to an imitative impairment are common in ASD. Also affected are core neuropsychological measures related to the sensorimotor integration (Hannant et al., 2016) in the auditory-vocal modality (O’Connor, 2012), such as processing speed. Genetics’s related contribution in this regard comes from the

CNTNAP2 which is closely related both with vocal learning (Alarcon et al., 2008; Condro & White, 2014) and potentiality of the risk to ASD (Peñagarikano & Geschwind, 2012). Even repetitive behavior, one of the hallmarks of ASD condition, which is found to be correlated to the aberrant development of the striatum, may be a by-product of deviant development of speech processing.

All in all, the current dissertation presents the overall hypothesis that vocal learning could serve as a foundational basis for language development (ontogeny) and evolution (phylogeny). We gather evidence from behavior, neuroanatomy, neurophysiology, genetics and evolution. By relating findings from these different domains, a number of hypotheses have been presented. Further research will help to single out those that are on the right track.

References

Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. *Trends in Neurosciences*, *31*(6), 265–272. https://doi.org/10.1016/j.tins.2008.02.011

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*(6), 529–546. https://doi.org/10.1017/S0140525X13003099

Ackermann, H., Mathiak, K., & Riecker, A. (2007). The contribution of the cerebellum to speech production and speech perception: Clinical and functional imaging data. *The Cerebellum*, *6*(3), 202–213. https://doi.org/10.1080/14734220701266742

Adam, I., Mendoza, E., Kobalz, U., Wohlgemuth, S., & Scharff, C. (2017). CNTNAP2 is a direct FoxP2 target in vitro and in vivo in zebra finches: complex regulation by age and activity. *Genes, Brain and Behavior*. Accepted manuscript online, DOI: 10.1111/gbb.12390

Alarcón, M., Abrahams, B. S., Stone, J. L., Duvall, J. A., Perederiy, J. V., Bomar, J. M., … Geschwind, D. H. (2008). Linkage, Association, and Gene-Expression Analyses Identify CNTNAP2 as an Autism-Susceptibility Gene. *American Journal of Human Genetics*, *82*(1), 150–159.

https://doi.org/10.1016/j.ajhg.2007.09.005

Alcock, K. J., Passingham, R. E., Watkins, K., & Vargha-Khadem, F. (2000). Pitch and Timing Abilities in Inherited Speech and Language Impairment. *Brain and Language*, *75*(1), 34–46. https://doi.org/10.1006/brln.2000.2323

Allen, T. A., & Fortin, N. J. (2013). The evolution of episodic memory. *Proceedings of the National Academy of Sciences*, *110*(Supplement_2), 10379–10386. https://doi.org/10.1073/pnas.1301199110

American Speech-Language-Hearing Association respectfully submits the enclosed report and recommendations for consideration. (2012). Asha'S Recommended Revisions To the Dsm-5. Social Sciences, (June).

American Psychiatric Association. (2003). APA (2000). *Diagnostic and statistical manual of mental disorders*, *4*.

Apfelbach, R. (1972). Electrically elicited vocalizations in the gibbon Hylobates lar (Hylobatidae), and their behavioral significance. *Ethology*, *30*(4), 420-430.

Arriaga, G., Zhou, E. P., & Jarvis, E. D. (2012). Of Mice, Birds, and Men: The Mouse Ultrasonic Song System Has Some Features Similar to Humans and Song-Learning Birds. *PLoS ONE*, *7*(10). https://doi.org/10.1371/journal.pone.0046610

Atoji, Y., & Wild, J. M. (2006). Anatomy of the Avian Hippocampal Formation. *Reviews in the Neurosciences*, *17*(1–2), 3–16. https://doi.org/10.1515/REVNEURO.2006.17.1-2.3

Bailey, D. J., & Wade, J. (2005). FOS and ZENK responses in 45-day-old zebra finches vary with auditory stimulus and brain region, but not sex. *Behavioural Brain Research*, *162*(1), 108–115. https://doi.org/10.1016/j.bbr.2005.03.016

Bakkaloglu, B., O'Roak, B. J., Louvi, A., Gupta, A. R., Abelson, J. F., Morgan, T. M., … State, M. W. (2008). Molecular Cytogenetic Analysis and Resequencing of Contactin Associated Protein-Like 2 in Autism Spectrum Disorders. *American Journal of Human Genetics*, *82*(1), 165–173. https://doi.org/10.1016/j.ajhg.2007.09.017

Balari, S., & Lorenzo, G. (2015). It is an organ, it is new, but it is not a new organ. Conceptualizing language from a homological perspective. *Frontiers in Ecology and Evolution*, *3*, 58. https://doi.org/10.3389/fevo.2015.00058

Baptista, Luis F.; Gaunt, Sandra L. L. (1997). Social interaction and vocal development in birds. In Snowdon, Charles T. (Ed); Hausberger, Martine (Ed). (1997). *Social influences on vocal development*, (pp. 23-40). New York, NY, US: Cambridge University Press, ix, 352 pp. http://dx.doi.org/10.1017/CBO9780511758843.003

Barnea-Goraly, N., Kwon, H., Menon, V., Eliez, S., Lotspeich, L., & Reiss, A. L. (2004). White matter structure in autism: Preliminary evidence from diffusion tensor imaging. *Biological Psychiatry*, *55*(3), 323–326. https://doi.org/10.1016/j.biopsych.2003.10.022

Bartolotti, J., Bradley, K., Hernandez, A. E., & Marian, V. (2017). Neural signatures of second language learning and control. *Neuropsychologia*, *98*, 130–138. https://doi.org/10.1016/j.neuropsychologia.2016.04.007

Bass, A. H., Gilland, E. H., & Baker, R. (2008). Evolutionary Origins for Social Vocalisation in a Vertebrate Hindbrain-Spinal Compartment. *Science*, *321*(5887), 417–421. https://doi.org/10.1126/science.1157632.Evolutionary

Bebko, J. M., Schroeder, J. H., & Weiss, J. A. (2014). The McGurk effect in children with autism and asperger syndrome. *Autism Research*, *7*(1), 50–59. https://doi.org/10.1002/aur.1343

Behbehani, M. M. (1995). Functional characteristics of the midbrain periaqueductal gray. *Progress in Neurobiology*, *46*(6), 575–605. https://doi.org/10.1016/0301-0082(95)00009-K

Benarroch, E. E. (2012). Periaqueductal gray: An interface for behavioral control. *Neurology*, *78*(3), 210–217. https://doi.org/10.1212/WNL.0b013e31823fcdee

Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences*, *15*(3), 113–121. https://doi.org/10.1016/j.tics.2011.01.002

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences, 17(2)*, 89–98. http://doi.org/10.1016/j.tics.2012.12.002

Berwick, R. C., & Chomsky, N. (2015). *Why only us: Language and evolution*. MIT press.

Bicanic, I., Bornschein, U., Enard, W., Hevers, W., Paabo, S., & Petanjek, Z. (2014). Regional differences in dendritic morphology of medium spiny striatal neurons in Foxp2 mice are influenced by substitution at position T302N. In *9th FENS*

Forum of Neuroscience.

Boddaert, N., Chabane, N., Gervais, H., Good, C. D., Bourgeois, M., Plumet, M. H., … Zilbovicius, M. (2004). Superior temporal sulcus anatomical abnormalities in childhood autism: A voxel-based morphometry MRI study. *NeuroImage*, *23*(1), 364–369. https://doi.org/10.1016/j.neuroimage.2004.06.016

Boeckx, C. (2016). A conjecture about the neural basis of recursion in light of descent with modification. *Journal of Neurolinguistics*, 6–11. https://doi.org/10.1016/j.jneuroling.2016.08.003

Boeckx, C., Grohmann, K. K., & Others. (2007). Biolinguistics *1:1*, 1–8.

Bolhuis, J. J., & Gahr, M. (2006). Neural mechanisms of birdsong memory. *Nature Reviews Neuroscience*, *7*(5), 347–357. https://doi.org/10.1038/nrn1904

Bolhuis, J. J., Tattersall, I., Chomsky, N., & Berwick, R. C. (2014). How Could Language Have Evolved? *PLoS Biology, 12(8),* 1–6. https://doi.org/10.1371/journal.pbio.1001934

Bond, A. M., VanGompel, M. J. W., Sametsky, E. A., Clark, M. F., Savage, J. C., Disterhoft, J. F., & Kohtz, J. D. (2009). Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nature Neuroscience*, *12*(8), 1020–1027. https://doi.org/10.1038/nn.2371

Bond, J., & Woods, C. G. (2006). Cytoskeletal genes regulating brain size. *Current Opinion in Cell Biology*, *18*(1), 95–101. https://doi.org/10.1016/j.ceb.2005.11.004

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A.-M. (2003).

Enhanced Pitch Sensitivity in Individuals with Autism: A Signal Detection Analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226–235. https://doi.org/10.1162/089892903321208169

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., & Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: Common computational properties. *Trends in Cognitive Sciences*, *19*(3), 142–150. https://doi.org/10.1016/j.tics.2014.12.008

Bowers, J. S., Mattys, S. L., & Gage, S. H. (2009). Preserved Implicit Knowledge of a Forgotten Childhood Language, 1–6.

Bradford-Heit, a., & Dodd, B. (1998). Learning new words using imitation and additional cues: differences between children with disordered speech. *Child Language Teaching and Therapy*, *14*(2), 159–179. https://doi.org/10.1191/026565998674298897

Brauer, J., Anwander, A., & Friederici, A. D. (2011). Neuroanatomical prerequisites for language functions in the maturing brain. *Cerebral Cortex*, *21*(2), 459–466. https://doi.org/10.1093/cercor/bhq108

Brenowitz, E. A., & Beecher, M. D. (2005). Song learning in birds: Diversity and plasticity, opportunities and challenges. *Trends in Neurosciences*, *28*(3), 127–132. https://doi.org/10.1016/j.tins.2005.01.004

Brown, T. G. (1915). Note on the physiology of the basal ganglia and mid-brain of the anthropoid ape, especially in reference to the act of laughter. *The Journal of physiology*, *49*(4), 195.

Calderoni, S., Bellani, M., Hardan, A., Muratori, F., & Brambilla, P. (2014). Basal ganglia and restricted and repetitive behaviours in Autism Spectrum Disorders: Current status and future perspectives. *Epidemiology and Psychiatric Sciences*, *23*(May), 235–238. https://doi.org/http://dx.doi.org/10.1017/S2045796014000171

Camacho-Schlenker, S., Courvoisier, H., & Aubin, T. (2011). Song sharing and singing strategies in the winter wren troglodytes. *Behavioural Processes*, *87*(3), 260–267. https://doi.org/10.1016/j.beproc.2011.05.003

Campbell, P., Reep, R. L., Stoll, M. L., Ophir, A. G., & Phelps, S. M. (2009). Conservation and diversity of Foxp2 expression in muroid rodents: Functional implications. *Journal of Comparative Neurology*, *512*(1), 84–100. https://doi.org/10.1002/cne.21881

Cappe, C., Rouiller, E. M., & Barone, P. (2009). Multisensory anatomical pathways. *Hearing Research*, *258*(1–2), 28–36. https://doi.org/10.1016/j.heares.2009.04.017

Carmen Panaitof, S., Abrahams, B. S., Dong, H., Geschwind, D. H., & White, S. A. (2010). Language-related cntnap2 gene is differentially expressed in sexually dimorphic song nuclei essential for vocal learning in songbirds. *Journal of Comparative Neurology*, *518*(11), 1995–2018. https://doi.org/10.1002/cne.22318

Catani, M., Jones, D. K., & Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, *57*(1), 8–16. https://doi.org/10.1002/ana.20319

Catmur, C., Walsh, V., & Heyes, C. (2009). Associative sequence learning: the role of experience in the development of imitation and the mirror system. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1528)*, 2369–2380. https://doi.org/10.1098/rstb.2009.0048

Chabout, J., Sarkar, A., Patel, S. R., Radden, T., Dunson, D. B., Fisher, S. E., & Jarvis, E. D. (2016). A Foxp2 mutation implicated in human speech deficits alters sequencing of ultrasonic vocalizations in adult male mice. *Frontiers in Behavioral Neuroscience*, *10*(October), 197. https://doi.org/10.3389/FNBEH.2016.00197

Chakraborty, M., & Jarvis, E. D. (2015). Brain evolution by brain pathway duplication. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1684), 20150056. https://doi.org/10.1098/rstb.2015.0056

Chakraborty, M., Walløe, S., Nedergaard, S., Fridel, E. E., Dabelsteen, T., Pakkenberg, B., … Jarvis, E. D. (2015). Core and shell song systems unique to the parrot brain. *PLoS ONE*, *10*(6), 1–37. https://doi.org/10.1371/journal.pone.0118496

Chen, Q., Heston, J. B., Burkett, Z. D., & White, S. A. (2013). Expression analysis of the speech-related genes FoxP1 and FoxP2 and their relation to singing behavior in two songbird species. *Journal of Experimental Biology*, *216*(19), 3682–3692. https://doi.org/10.1242/jeb.085886

Choi, J., Cutler, A., & Broersma, M. (2017). Early development of abstract language knowledge: evidence from perception–production transfer of birth-language memory. *Royal Society Open Science*, *4*(1). 160660.

Cociu, B. A., Das, S., Billeci, L., Jamal, W., Maharatna, K., Calderoni, S., ... & Muratori, F. (2017). Multimodal Functional and Structural Brain Connectivity Analysis in Autism: A Preliminary Integrated Approach with EEG, fMRI and DTI. *IEEE Transactions on Cognitive and Developmental Systems*. PP (99), 10.1109/TCDS.2017.2680408

Condro, M. C., & White, S. A. (2014). Distribution of language-related Cntnap2 protein in neural circuits critical for vocal learning. *Journal of Comparative Neurology*, *522*(1), 169–185. https://doi.org/10.1002/cne.23394

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, *5*(12), 539–546. https://doi.org/10.1016/S1364-6613(00)01800-3

Corballis, M. C. (1999). Did Language Evolve before Speech ? *The Evolution of Human Language*, (1987), 115–123. https://doi.org/10.1017/CBO9780511817755.008

Corballis, M. C. (2009). The evolution of language. *Annals of the New York Academy of Sciences*, *1156*, 19–43. https://doi.org/10.1111/j.1749-6632.2009.04423.x

Curtiss, S. (1977). *Genie: a psycholinguistic study of a modern-day wild child.* Academic Press. NY.

Cynx, J. (1995). Similarities in absolute and relative pitch perception in songbirds (starling and zebra finch) and a nonsongbird (pigeon). *Journal of Comparative Psychology*, *109*(3), 261–267. https://doi.org/10.1037/0735-7036.109.3.261

Dalziell, A. H., Peters, R. A., Cockburn, A., Dorland, A. D., Maisey, A. C., & Magrath, R. D. (2013). Dance choreography is coordinated with song repertoire in a complex avian display. *Current Biology*, *23*(12), 1132–1135. https://doi.org/10.1016/j.cub.2013.05.018

DasGupta, S., Ferreira, C. H., & Miesenbock, G. (2014). FoxP influences the speed and accuracy of a perceptual decision in Drosophila. *Science*, *344*(6186), 901–904. https://doi.org/10.1126/science.1252114

Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: The role of prediction and context. *Trends in Cognitive Sciences*, *19*(2), 92–99. https://doi.org/10.1016/j.tics.2014.12.004

Deacon, T. W. (1989). The neural circuitry underlying primate calls and human language. *Human Evolution*, *4*(5), 367–401. https://doi.org/10.1007/BF02436435

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences*, *95*(24), 14529–14534. https://doi.org/10.1073/pnas.95.24.14529

Devoogd, T. J., Krebs, J. R., Healy, S. D., & Purvis, A. (1993). Relations between Song Repertoire Size and the Volume of Brain Nuclei Related to Song: Comparative Evolutionary Analyses amongst Oscine Birds. *Proceedings of the Royal Society B: Biological Sciences*, *254*(1340), 75–82. https://doi.org/10.1098/rspb.1993.0129

Diller, K. C., & Cann, R. L. (2009). Evidence against a genetic-based revolution in language 50,000 years ago. *The Cradle of Language*, 135–149.

Drew, L. J., Fusi, S., & Hen, R. (2013). Adult neurogenesis in the mammalian hippocampus: Why the dentate gyrus? *Learning & Memory*, *20*(12), 710–729. https://doi.org/10.1101/lm.026542.112

Dufour, V., Poulin, N., Charlotte Curé, & Sterck, E. H. M. (2015). Chimpanzee drumming: a spontaneous performance with characteristics of human musical drumming. *Scientific Reports*, *5*, 11320. https://doi.org/10.1038/srep11320

Dutschmann, M., Mörschel, M., Kron, M., & Herbert, H. (2004). Development of adaptive behaviour of the respiratory network: Implications for the pontine Kölliker-Fuse nucleus. *Respiratory Physiology and Neurobiology*, *143*(2–3), 155–165. https://doi.org/10.1016/j.resp.2004.04.015

Dutschmann, M. (2016). Exploring Degeneracy of Brainstem Networks That Generate the Respiratory Rhythm. Experimental Biology meeting. San Diego.

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*(1), 109–120. https://doi.org/10.1016/j.neuron.2004.08.028

Elizabeth Redcay, & Eric Courchesne. (2008). Deviant fMRI patterns of brain activity to speech in 2–3 year-old children with autism spectrum disorder. *Biological Psychiatry*, *64*(7), 589–598. https://doi.org/10.1038/jid.2014.371

Enard, W. (2016). The Molecular Basis of Human Brain Evolution. *Current Biology*, *26*(20), R1109–R1117. https://doi.org/10.1016/j.cub.2016.09.030

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S. M., Blass, T., Somel, M., … Pääbo, S. (2009). A Humanized Version of Foxp2 Affects Cortico-Basal Ganglia

Circuits in Mice. *Cell*, *137*(5), 961–971. https://doi.org/10.1016/j.cell.2009.03.041

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., … Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, *418*(6900), 869–872. https://doi.org/10.1038/nature01025

Esposito, A., Demeurisse, G., Alberti, B., & Fabbro, F. (1999). Complete mutism after midbrain periaqueductal gray lesion. *Neuroreport*, *10*(4), 681–5. https://doi.org/10.1097/00001756-199903170-00004

Evans, K. E., & Demuth, K. (2012). Individual differences in pronoun reversal: Evidence from two longitudinal case studies. *Journal of child language*, *39*(01), 162-191.

Evans, P. D. (2005). Microcephalin, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans. *Science*, *309*(5741), 1717–1720. https://doi.org/10.1126/science.1113722

Evans, P. D., Vallender, E. J., & Lahn, B. T. (2006). Molecular evolution of the brain size regulator genes CDK5RAP2 and CENPJ. *Gene*, *375*(1–2), 75–79. https://doi.org/10.1016/j.gene.2006.02.019

Everaert, M. B., Huybregts, M. A., Berwick, R. C., Chomsky, N., Tattersall, I., Moro, A., & Bolhuis, J. J. (2017). What is language and how could it have evolved?. *Trends in Cognitive Sciences*.

Ewing, A. W. (1983). Functional Aspects of Drosophila Courtship. *Biological Reviews*, *58*(2), 275–292. https://doi.org/10.1111/j.1469-185X.1983.tb00390.x

Eyler, L. T., Pierce, K., & Courchesne, E. (2012). A failure of left temporal cortex to specialize for language is an early emerging and fundamental property of autism. *Brain*, *135*(3), 949–960. https://doi.org/10.1093/brain/awr364

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, (2004), 491–541. https://doi.org/10.1017/S0140525X04000111

Fee, M. S., & Goldberg, J. H. (2011). A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience*, *198*, 152–170. https://doi.org/10.1016/j.neuroscience.2011.09.069

Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E., … Jarvis, E. D. (2008). Molecular mapping of movement-associated areas in the avian brain: A motor theory for vocal learning origin. *PLoS ONE*, *3*(3). https://doi.org/10.1371/journal.pone.0001768

Finney, E. M. F. (2001). Visual stimuli activate auditory cortex in the deaf. *Nature Neuroscience*, *4*(12), 1171. https://doi.org/10.1038/nn763

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*(7), 258–267. https://doi.org/10.1016/S1364-6613(00)01494-7

Fitch, W. T. (2009). Fossil cues to the evolution of speech. In *The cradle of language*, ed.by Botha, R. & Knight, C. pp 112-134. Oxford University Press.

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

Fitch, W. T. (2011). The evolution of syntax: An exaptationist perspective. *Frontiers in Evolutionary Neuroscience*, *3*(DEC), 1–12. https://doi.org/10.3389/fnevo.2011.00009

Fitch, W. T., & Jarvis, E. D. (2013). Birdsong and other animal models for human speech, song, and vocal learning. In *Language, music and the brain: a Mysterious Relationship*, ed. by Arbib, M. A. pp 499-540. The MIT Press, CA.

Fitch, W. T. (2015). Four principles of bio-musicology. *Phil. Trans. R. Soc. B*, *370*(1664), 20140091.

Fitch, W. T. (2017). Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-017-1236-5

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.

Ford, A. A., Triplett, W., Sudhyadhom, A., Gullett, J., McGregor, K., FitzGerald, D. B., … Crosson, B. (2013). Broca's area and its striatal and thalamic connections: a diffusion-MRI tractography study. *Frontiers in Neuroanatomy*, *7*(May), 1–12. https://doi.org/10.3389/fnana.2013.00008

Frey, U., Ltp, R. G. M. M. L., Rizzolatti, G., Arbib, M. A., & Rizzolatti, G. (1998). Rizzolati-Arbib, *2236*(1988), 1667–1669. https://doi.org/10.1016/S0166-2236(98)01260-0

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16*(5), 262–268. https://doi.org/10.1016/j.tics.2012.04.001

Fritz, J., Elhilali, M., & Shamma, S. (2005). Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hearing Research*, *206*(1–2), 159–176. https://doi.org/10.1016/j.heares.2005.01.015

Fujita, E., Tanabe, Y., Shiota, A., Ueda, M., Suwa, K., Momoi, M. Y., & Momoi, T. (2008). Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(8), 3117–22. https://doi.org/10.1073/pnas.0712298105

Gage, N. M., Juranek, J., Filipek, P. A., Osann, K., Flodman, P., Isenberg, A. L., & Spence, M. A. (2009). Rightward hemispheric asymmetries in auditory language cortex in children with autistic disorder: An MRI investigation. *Journal of Neurodevelopmental Disorders*, *1*(3), 205–214. https://doi.org/10.1007/s11689-009-9010-2

Gahr, M. (2000). Neural song control system of hummingbirds: Comparison to swifts, vocal learning (songbirds) and nonlearning (suboscines) passerines, and vocal learning (budgerigars) and nonlearning (dove, owl, gull, quail, chicken) nonpasserines. *Journal of Comparative Neurology*, *426*(2), 182–196. https://doi.org/10.1002/1096-9861(20001016)426:2<182::AID-CNE2>3.0.CO;2-M

Ganz, J., Kaslin, J., Freudenreich, D., Machate, A., Geffarth, M., & Brand, M. (2011). Subdivisions of the adult zebrafish subpallium by molecular marker analysis Julia Ganz. *Journal of Comparative Neurology*, *520*(Sfb 655), 633–655. https://doi.org/10.1002/cne.

Gernsbacher, M. A., Morson, E. M., & Grace, E. J. (2016). Language and speech in autism. *Annual Review of Linguistics*, *2*, 413-425.

Genty, E., Clay, Z., Hobaiter, C., & Zuberbühler, K. (2014). Multi-modal use of a socially directed call in bonobos. *PloS one*, *9*(1), e84738.

Glazebrook, C., Gonzalez, D., Hansen, S., & Elliott, D. (2009). The role of vision for online control of manual aiming movements in persons with autism spectrum disorders. *Autism*, *13*(4), 411–433. https://doi.org/10.1177/1362361309105659

Gomot, M., Belmonte, M. K., Bullmore, E. T., Bernard, F. A., & Baron-Cohen, S. (2008). Brain hyper-reactivity to auditory novel targets in children with high-functioning autism. *Brain*, *131*(9), 2479–2488. https://doi.org/10.1093/brain/awn172

Goucha, T., Zaccarella, E., & Friederici, A. D. (2017). A revival of the Homo loquens as a builder of labeled structures: neurocognitive considerations. *Neuroscience & Biobehavioral Reviews*, 1–12. https://doi.org/10.1016/j.neubiorev.2017.01.036

Gray, P. A. (2008). Transcription factors and the genetic organization of brain stem respiratory neurons, *Journal of Applied Physiology*, *104*(5), 1513-1521. https://doi.org/10.1152/japplphysiol.01383.2007.

Graybiel, A. M. (2005). The basal ganglia: Learning new tricks and loving it. *Current*

Opinion in Neurobiology, *15*(6), 638–644. https://doi.org/10.1016/j.conb.2005.10.006

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Paabo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, *14*(4), 531–551. https://doi.org/10.1017/S0140525X00071235

Groen, W. B., Van Orsouw, L., Huurne, N. Ter, Swinkels, S., Van Der Gaag, R. J., Buitelaar, J. K., & Zwiers, M. P. (2009). Intact spectral but abnormal temporal processing of auditory stimuli in autism. *Journal of Autism and Developmental Disorders*, *39*(5), 742–750. https://doi.org/10.1007/s10803-008-0682-3

Groenewegen, H. J. (2003). The Basal Ganglia and Motor Control. *Neural Plasticity*, *10*(1–2), 107–120. https://doi.org/10.1155/NP.2003.107

Groszer, M., Keays, D. A., Deacon, R. M. J., de Bono, J. P., Prasad-Mulcare, S., Gaub, S., … Fisher, S. E. (2008). Impaired Synaptic Plasticity and Motor Learning in Mice with a Point Mutation Implicated in Human Speech Deficits. *Current Biology*, *18*(5), 354–362. https://doi.org/10.1016/j.cub.2008.01.060

Haesler, S. (2004). FoxP2 Expression in Avian Vocal Learners and Non-Learners. *Journal of Neuroscience*, *24*(13), 3164–3175. https://doi.org/10.1523/JNEUROSCI.4369-03.2004

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), 2885–2897. https://doi.org/10.1371/journal.pbio.0050321

Hamaguchi, K., & Mooney, R. (2012). Recurrent Interactions between the Input and Output of a Songbird Cortico-Basal Ganglia Pathway Are Implicated in Vocal Sequence Variability. *Journal of Neuroscience*, *32*(34), 11671–11687. https://doi.org/10.1523/JNEUROSCI.1666-12.2012

Hamilton, B. L. (1973). Projections of the nuclei of the periaqueductal gray matter in the cat. *Journal of Comparative Neurology*, *152*(1), 45-57.

Hannant, P., Cassidy, S., Tavassoli, T., & Mann, F. (2016). Sensorimotor Difficulties Are Associated with the Severity of Autism Spectrum Conditions. *Frontiers in Integrative Neuroscience*, *10*(August), 1–14. https://doi.org/10.3389/fnint.2016.00028

Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of autism and developmental disorders*, *36*(1), 5-25.

Hauser, M. D. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., ... & Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in psychology*, *5*, 401.

Heaton, P., Williams, K., Cummins, O., & Happe, F. (2008). Autism and pitch processing splinter skills: A group and subgroup analysis. *Autism*, *12*(2), 203–219. https://doi.org/10.1177/1362361307085270

Hecht, E. E., Gutman, D. A., Khreisheh, N., Taylor, S. V., Kilner, J., Faisal, A. A., … Stout, D. (2014). Acquisition of Paleolithic toolmaking abilities involves structural remodeling to inferior frontoparietal regions. *Brain Structure and Function*, *220*(4), 2315–2331. https://doi.org/10.1007/s00429-014-0789-6

Hecht, E. E., Murphy, L. E., Gutman, D. A., Votaw, J. R., Schuster, D. M., Preuss, T. M., … Parr, L. A. (2013). Differences in Neural Activation for Object-Directed Grasping in Chimpanzees and Humans. *Journal of Neuroscience*, *33*(35), 14117–14134. https://doi.org/10.1523/JNEUROSCI.2172-13.2013

Heffner, Rickye S.; Heffner, H. E. (1982). Hearing in the elephant (Elephas maximus): Absolute sensitivity, frequency discrimination, and sound localization. *Journal of Comparative and Physiological Psychology*, *96*(6), 926–944.

Heinrich, R., Kunst, M., & Wirmer, A. (2012). Reproduction-related sound production of grasshoppers regulated by internal state and actual sensory environment. *Frontiers in Neuroscience*, *6*(JUN), 1–9. https://doi.org/10.3389/fnins.2012.00089

Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition*, *16*(2), 129–219. https://doi.org/10.1016/0010-0277(84)90003-9

Heyes, C. (2016). Homo imitans? Seven reasons why imitation couldn’t possibly be associative. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150069. https://doi.org/10.1098/rstb.2015.0069

Hickok, G. (2009). Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans. *Journal of Cognitive Neuroscience*, *21*(7), 1229–1243. https://doi.org/10.1162/jocn.2009.21189

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135-145.

Hickok, G. (2014). *The myth of mirror neurons: The real neuroscience of communication and cognition*. WW Norton & Company.

Hickok, G. (2016). A cortical circuit for voluntary laryngeal control: Implications for the evolution language. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-016-1100-z

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1–2), 67–99. https://doi.org/10.1016/j.cognition.2003.10.011

Higham, J. P., & Hebets, E. A. (2013). An introduction to multimodal communication. *Behavioral Ecology and Sociobiology*, *67*(9), 1381–1388. https://doi.org/10.1007/s00265-013-1590-x

Hill, E. L. (2001). Non-specific nature of specific language impairment: a review of the literature with regard to concomitant motor impairments. *International Journal of Language & Communication Disorders*, *36*(2), 149–171. https://doi.org/10.1080/13682820010019874

Hockett, C. (1960). The origin of speech. *Human Language and Animal Communication, 5-12.*

Hollander, E., Anagnostou, E., Chaplin, W., Esposito, K., Haznedar, M. M., Licalzi, E., … Buchsbaum, M. (2005). Striatal volume on magnetic resonance imaging and repetitive behaviors in autism. *Biological Psychiatry*, *58*(3), 226–232. https://doi.org/10.1016/j.biopsych.2005.03.040

Holstege, G., & Subramanian, H. H. (2016). Two different motor systems are needed to generate human speech. *Journal of Comparative Neurology*, *524*(8), 1558–1577. https://doi.org/10.1002/cne.23898

Honda, E., & Okanoya, K. (1999). Acoustical and Syntactical Comparisons between Songs of the White-backed Munia (Lonchura striata) and Its Domesticated Strain, the Bengalese Finch (Lonchura striata var. domestica). *Zoological Science*, *16*(2), 319–326. https://doi.org/10.2108/zsj.16.319

Huang, W., Zhou, Z., Asrar, S., Henkelman, M., Xie, W., & Jia, Z. (2011). p21-Activated Kinases 1 and 3 Control Brain Size through Coordinating Neuronal Complexity and Synaptic Properties. *Molecular and Cellular Biology*, *31*(3), 388–403. https://doi.org/10.1128/MCB.00969-10

Hubbard, A. L., Wilson, S. M., Callan, D. E., & Dapretto, M. (2009). Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping*, *30*(3), 1028–1037. https://doi.org/10.1002/hbm.20565

Hunt, G. R. (1996). Manufacture and use of hook-tools by New Caledonian crows. *Nature*, *379*(6562), 249–251. https://doi.org/10.1038/379249a0

Hunt, G. R., Rutledge, R. B., & Gray, R. D. (2006). The right tool for the job: What strategies do wild New Caledonian crows use? *Animal Cognition*, *9*(4), 307–316. https://doi.org/10.1007/s10071-006-0047-2

Hurford, J. R. (2004). Language Beyond Our Grasp: What Mirror Neurons Can, and Cannot, Do for Language Evolution. *Evolution of Communication Systems: A Comparative Approach*, *169*(2275), 297–314.

Jacob, F. (1977). Evolution and tinkering. *Science* 196: 4295, 1161-1166.

Jakobson, R. (1941). *Child Language, Aphasia and Linguistic Universals*. The Hague: Mouton.

Jarvis, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, *1016*(1), 749-777.

Jarvis, E. D. (2007). Neural systems for vocal learning in birds and humans: A synopsis. *Journal of Ornithology*, *148*(SUPPL. 1). https://doi.org/10.1007/s10336-007-0243-0

Jarvis, E. D. (2009). Bird Song Systems: Evolution. *Encyclopedia of Neuroscience*, *2*, 217–225.

Jarvis, E. D., Güntürkün, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., … Butler, A. B. (2005). Opinion: Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience*, *6*(2), 151–159. https://doi.org/10.1038/nrn1606

Jusczyk, P. W., & Luce, P. A. (1994). Infants′ Sensitivity to Phonotactic Patterns in the

Native Language. *Journal of Memory and Language*. https://doi.org/10.1006/jmla.1994.1030

Jürgens, U., & Ploog, D. (1970). Cerebral representation of vocalization in the squirrel monkey. *Experimental Brain Research*, *10*(5), 532-554.

Kern, J. K., Trivedi, M. H., Grannemann, B. D., Garver, C. R., Johnson, D. G., Andrews, A. A., … Schroeder, J. L. (2007). Sensory correlations in autism. *Autism*, *11*(2), 123–134. https://doi.org/10.1177/1362361307075702

Kittelberger, J. M., Land, B. R., & Bass, A. H. (2006). Midbrain periaqueductal gray and vocal patterning in a teleost fish. *Journal of neurophysiology*, *96*(1), 71-85.

Kittelberger, J. M., & Bass, A. H. (2013). Vocal-motor and auditory connectivity of the midbrain periaqueductal gray in a teleost fish. *Journal of Comparative Neurology*, *521*(4), 791–812. https://doi.org/10.1002/cne.23202

Kojima, S. (1990). Comparison of Auditory Functions in the Chimpanzee and Human. *Folia Primatol*, *55*, 62–72.

Kojima, S. (2003). *A Search for the Origins of Human Speech: Auditory and vocal functions of the chimpanzee*. Kyoto University Press and Trans Pacific Press, Kyoto, Japan.

Konopka, G., Bomar, J. M., Winden, K., Coppola, G., Jonsson, Z. O., Gao, F., … Geschwind, D. H. (2009). Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature*, *462*(7270), 213–217. https://doi.org/10.1038/nature08549

Kraus, K. S., & Canlon, B. (2012). Neuronal connectivity and interactions between the auditory and limbic systems. Effects of noise and tinnitus. *Hearing Research*, *288*(1–2), 34–46. https://doi.org/10.1016/j.heares.2012.02.009

Krause, J., Orlando, L., Serre, D., Viola, B., Prüfer, K., Richards, M. P., … Pääbo, S. (2007). Neanderthals in central Asia and Siberia. *Nature*, *449*(7164), 902–904. https://doi.org/10.1038/nature06193

Kriengwatana, B., Spierings, M. J., & ten Cate, C. (2016). Auditory discrimination learning in zebra finches: Effects of sex, early life conditions and stimulus characteristics. *Animal Behaviour*, *116*, 99–112. https://doi.org/10.1016/j.anbehav.2016.03.028

Kuhl, P. K. (1992). Psychoacoustics and speech perception: Internal standards, perceptual anchors, and prototypes. In *Developmental psychoacoustics*. Ed. by Werner, Lynne A. & Rubel, Edwin W. pp. 293-332. Washington, DC

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: Behavioral and electrophysiological measures. *Developmental Science*, *8*(1). https://doi.org/10.1111/j.1467-7687.2004.00384.x

Kuhl, P. K., Kuhl, P. K., Coffey-corina, S., Coffey-corina, S., Padden, D., Padden, D., … Dawson, G. (2005). Links between social and linguistic processing of speech in children with autism: behavioural and electrophysiological measures. *Developmental Science*, *8*(1), 1–12. https://doi.org/10.1111/j.1467-7687.2004.00384.x

Kumar, S., Bonnici, H. M., Teki, S., Agus, T. R., Pressnitzer, D., Maguire, E. A., &

Griffiths, T. D. (2014). Representations of specific acoustic patterns in the auditory cortex and hippocampus. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1791), 20141000. https://doi.org/10.1098/rspb.2014.1000

Kumar, V., Croxson, P. L., & Simonyan, K. (2016). Structural Organization of the Laryngeal Motor Cortical Network and Its Implication for Evolution of Speech Production. *Journal of Neuroscience*, *36*(15), 4170–4181. https://doi.org/10.1523/JNEUROSCI.3914-15.2016

Lai, C. S. L., 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. https://doi.org/10.1038/35097076

Lai, G., Schneider, H. D., Schwarzenberger, J. C., & Hirsch, J. (2011). Speech Stimulation during Functional MR Imaging as a Potential Indicator of Autism. *Radiology*, *260*(2), 521–530. https://doi.org/10.1148/radiol.11101576

Laland, K. (2016). Life's Intimate Dance. *Trends in Ecology & Evolution*, *31*(12), 889–890. https://doi.org/10.1016/j.tree.2016.09.003

Lambert, M. L., Seed, A. M., & Slocombe, K. E. (2015). A novel form of spontaneous tool use displayed by several captive greater vasa parrots (Coracopsis vasa): Table 1. *Biology Letters*, *11*(12), 20150861. https://doi.org/10.1098/rsbl.2015.0861

Langen, M., Schnack, H. G., Nederveen, H., Bos, D., Lahuis, B. E., de Jonge, M. V., … Durston, S. (2009). Changes in the Developmental Trajectories of Striatum in Autism. *Biological Psychiatry*, *66*(4), 327–333. https://doi.org/10.1016/j.biopsych.2009.03.017

Lawton, K. J., Wassmer, T. L., & Deitcher, D. L. (2014). Conserved role of Drosophila melanogaster FoxP in motor coordination and courtship song. *Behavioural Brain Research*, *268*, 213–221. https://doi.org/10.1016/j.bbr.2014.04.009

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *Journal of Neuroscience*, *30*(22), 7604-7612.

Lenneberg, E. H., Freda G Rebelsky, & Irene a Nichols. (1965). The Vocalization of Infants Born to Deaf and to Hearing Parents. *Human Development*, *8*, 23–37.

Lerna, A., Esposito, D., Conson, M., Russo, L., & Massagli, A. (2012). Social–communicative effects of the Picture Exchange Communication System (PECS) in autism spectrum disorders. *International journal of language & communication disorders*, *47*(5), 609-617.

Lewis, G. J., & Bates, T. C. (2013). The long reach of the gene. *Psychologist*, *26*(3), 194–198. https://doi.org/10.1162/jocn

Li, G., Wang, J., Rossiter, S. J., Jones, G., & Zhang, S. (2007). Accelerated FoxP2 evolution in echolocating bats. *PLoS ONE*, *2*(9). https://doi.org/10.1371/journal.pone.0000900

Liebal, K., Pika, S., & Tomasello, M. (2004). Social communication in siamangs (Symphalangus syndactylus): Use of gestures and facial expressions. *Primates*, *45*(1), 41–57. https://doi.org/10.1007/s10329-003-0063-7

Lieberman, P. (2016). The evolution of language and thought. *Journal of Anthropological Sciences*, *94*, 127–146. https://doi.org/10.4436/jass.94029

Liégeois, F. J., Hildebrand, M. S., Bonthrone, A., Turner, S. J., Scheffer, I. E., Bahlo, M., … Morgan, A. T. (2016). Early neuroimaging markers of FOXP2 intragenic deletion. *Scientific Reports*, *6*(September), 35192. https://doi.org/10.1038/srep35192

Lin, M., Pedrosa, E., Shah, A., Hrabovsky, A., Maqbool, S., Zheng, D., & Lachman, H. M. (2011). RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS ONE*, *6*(9). https://doi.org/10.1371/journal.pone.0023356

Liu, W., Wada, K., Jarvis, E., & Nottebohm, F. (2013). Rudimentary substrates for vocal learning in a suboscine. *Nature Communications*, *4*(May), 1–12. https://doi.org/10.1038/ncomms3082

Lohr, B., Dooling, R. J., & Bartone, S. (2006). The discrimination of temporal fine structure in call-like harmonic sounds by birds. *Journal of Comparative Psychology*, *120*(3), 239–251. https://doi.org/10.1037/0735-7036.120.3.239

Lund, J. P., & Kolta, A. (2006). Brainstem circuits that control mastication: Do they have anything to say during speech? *Journal of Communication Disorders*, *39*(5), 381–390. https://doi.org/10.1016/j.jcomdis.2006.06.014

MacLarnon, A., & Hewitt, G. (2004). Increased breathing control: Another factor in the evolution of human language. *Evolutionary Anthropology*, *13*(5), 181–197.

https://doi.org/10.1002/evan.20032

Manuscript, A. (2008). NIH Public Access. *Growth (Lakeland)*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Maricic, T., Günther, V., Georgiev, O., Gehre, S., Ćurlin, M., Schreiweis, C., … Pääbo, S. (2013). A recent evolutionary change affects a regulatory element in the human FOXP2 gene. *Molecular Biology and Evolution*, *30*(4), 844–852. https://doi.org/10.1093/molbev/mss271

Marler, P., & Peters, S. (1988). The Role of Song Phonology and Syntax in Vocal Learning Preferences in the Song Sparrow, Melospiza melodia. *Ethology*, *77*(2), 125–149. https://doi.org/10.1111/j.1439-0310.1988.tb00198.x

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, *63*(1), 240–244. https://doi.org/10.1016/j.neuroimage.2012.06.043

Martin, J. R. (1976). Motivated behaviors elicited from hypothalamus, midbrain, and pons of the guinea pig (Cavia porcellus). *Journal of comparative and physiological psychology*, *90*(11), 1011 -1034. http://dx.doi.org/10.1037

Maseko, B. C., Spocter, M. A., Haagensen, M., & Manger, P. R. (2012). Elephants Have Relatively the Largest Cerebellum Size of Mammals. *Anatomical Record*, *295*(4), 661–672. https://doi.org/10.1002/ar.22425

Matsunaga, E., & Okanoya, K. (2014). Cadherins: Potential regulators in the faculty of language. *Current Opinion in Neurobiology*, *28*, 28–33.

https://doi.org/10.1016/j.conb.2014.06.001

Mayberry, R., & Jaques, J. (2000). Gesture productino during stuttered speech: insights into the nature of speech-gesture integration. *Language and Gesture*, 199-215.

McEvilly, R. J., de Diaz, M. O., Schonemann, M. D., Hooshmand, F., & Rosenfeld, M. G. (2002). Transcriptional regulation of cortical neuron migration by POU domain factors. *Science (New York, N.Y.)*, *295*(5559), 1528–32. https://doi.org/10.1126/science.1067132

Meltzoff, A. N., & Moore, M. K. (1994). Imitation, memory, and the representation of persons. *Infant behavior and development*, *17*(1), 83-99.

Mendoza, E., Colomb, J., Rybak, J., Pfluger, H. J., Zars, T., Scharff, C., & Brembs, B. (2014). Drosophila FoxP mutants are deficient in operant self-learning. *PLoS ONE*, *9*(6). https://doi.org/10.1371/journal.pone.0100648

Merker, B. (2012). The vocal learning constellation: Imitation, ritual culture, encephalization. In *Music, language and human evolution*, ed. by Bannan, N. pp 215-60. Oxford University Press.

Micheletta, J., Engelhardt, A., Matthews, L., Agil, M., & Waller, B. M. (2013). Multicomponent and multimodal lipsmacking in crested macaques (Macaca nigra). *American Journal of Primatology*, *75*(7), 763–773. https://doi.org/10.1002/ajp.22105

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford

University Press.

Moore, J. K. (2002). Maturation of human auditory cortex: implications for speech perception. *The Annals of Otology, Rhinology & Laryngology. Supplement*, *189*, 7–10. https://doi.org/10.1177/00034894021110S502

Moorman, S., Gobes, S. M. H., Kuijpers, M., Kerkhofs, A., Zandbergen, M. A., & Bolhuis, J. J. (2012). Human-like brain hemispheric dominance in birdsong learning. *Proceedings of the National Academy of Sciences*, *109*(31), 12782–12787. https://doi.org/10.1073/pnas.1207207109

Morley, I. (2014). A Multi-Disciplinary approach to the origins of music: Perspectives from anthropology, archaeology, cognition and behaviour. *Journal of Anthropological Sciences*, *92*(2014), 147–177. https://doi.org/10.4436/JASS.92008

Mühleisen, T. W., Leber, M., Schulze, T. G., Strohmaier, J., Degenhardt, F., Treutlein, J., … Cichon, S. (2014). Genome-wide association study reveals two new risk loci for bipolar disorder. *Nature Communications*, *5, 3339.* https://doi.org/10.1038/ncomms4339

Murakami, T., Kell, C. A., Restle, J., Ugawa, Y., & Ziemann, U. (2015). Left Dorsal Speech Stream Components and Their Contribution to Phonological Processing. *Journal of Neuroscience*, *35*(4), 1411–1422. https://doi.org/10.1523/JNEUROSCI.0246-14.2015

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage*, *59*(1), 781–787. https://doi.org/10.1016/j.neuroimage.2011.07.024

Nath, A. R., & Beauchamp, M. S. (2013). NIH Public Access. *NeuroImage*, *59*(1), 781–787. https://doi.org/10.1016/j.neuroimage.2011.07.024.A

Newbury, D. F., Bonora, E., Lamb, J. A., Fisher, S. E., Lai, C. S. L., Baird, G., … Merricks, M. J. (2002). *FOXP2* Is Not a Major Susceptibility Gene for Autism or Specific Language Impairment. *The American Journal of Human Genetics*, *70*(5), 1318–1327. https://doi.org/10.1086/339931

Nishimura, H., Hashikawa, K., Doi, K., Iwaki, T., Watanabe, Y., Kusuoka, H., … Kubo, T. (1999). Sign language "heard" in the auditory cortex. *Nature*, *397*(6715), 116. https://doi.org/10.1038/16376

Noens, I. L. J., & van Berckelaer-Onnes, I. A. (2008). The central coherence account of autism revisited: Evidence from the ComFor study. *Research in Autism Spectrum Disorders*, 2(2), 209–222. http://doi.org/10.1016/j.rasd.2007.05.004

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646. https://doi.org/10.1037/0033-295X.110.4.611

Nudel, R., & Newbury, D. F. (2013). Foxp2. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(5), 547–560. https://doi.org/10.1002/wcs.1247

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience and Biobehavioral Reviews*, *36*(2), 836–854. https://doi.org/10.1016/j.neubiorev.2011.11.008

Odom, K. J., Hall, M. L., Riebel, K., Omland, K. E., & Langmore, N. E. (2014). Female song is widespread and ancestral in songbirds. *Nature Communications*, *5*, 1–6. https://doi.org/10.1038/ncomms4379

Oliveras-Rentas, R. E., Kenworthy, L., Roberson, R. B., Martin, A., & Wallace, G. L. (2012). WISC-IV profile in high-functioning autism spectrum disorders: impaired processing speed is associated with increased autism communication symptoms and decreased adaptive communication abilities. *Journal of autism and developmental disorders*, *42*(5), 655-664.

Olkowicz, S., Kocourek, M., Lučan, R. K., Porteš, M., Fitch, W. T., Herculano-Houzel, S., & Němec, P. (2016). Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences*, *113*(26), 7255–7260. https://doi.org/10.1073/pnas.1517131113

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child development*, (59):2, 441-449.

Oller, D. K. (2014). *The emergence of the speech capacity*. Psychology Press.

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., … Slaughter, V. (2016). Comprehensive Longitudinal Study Challenges the Existence of Neonatal Imitation in Humans. *Current Biology*, *26*(10), 1–5. https://doi.org/10.1016/j.cub.2016.03.047

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*(51), 22026-22031.

Overath, T., Mcdermott, J. H., Zarate, J. M., & Poeppel, D. (2016). The coritcal analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nature Neuroscience*, *18*(6), 903–911. https://doi.org/10.1038/nn.4021.

Parey, V. P., & Hamburg, B. u. (1972). Electrically Elicited Vocalizations in the Gibbon Hylobates lar (Hylobatidae), and their Behavioral Significance. *Souderdruck Aus Zeitschrift Fur Tierpsycholgie*, *30*, 420–430.

Pariani, M. J., Spencer, A., Graham, J. M., & Rimoin, D. L. (2009). A 785 kb deletion of 3p14.1p13, including the FOXP1 gene, associated with speech delay, contractures, hypertonia and blepharophimosis. *European Journal of Medical Genetics*, *52*(2–3), 123–127. https://doi.org/10.1016/j.ejmg.2009.03.012

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception: An Interdisciplinary Journal*, *24*(1), 99-104.

Peeters, D., Snijders, T. M., Hagoort, P., & Ozyurek, A. (2017). Linking language to the visual world: Neural correlates of comprehending verbal reference to objects through pointing and visual cues. *Neuropsychologia*, *95*(December 2016), 21–29. https://doi.org/10.1016/j.neuropsychologia.2016.12.004

Pellicano, E., Gibson, L., Maybery, M., Durkin, K., & Badcock, D. R. (2005). Abnormal global processing along the dorsal visual pathway in autism: A possible mechanism for weak visuospatial coherence? *Neuropsychologia*, *43*(7), 1044–1053. https://doi.org/10.1016/j.neuropsychologia.2004.10.003

Peñagarikano, O., & Geschwind, D. H. (2012). What does CNTNAP2 reveal about autism spectrum disorder? *Trends in Molecular Medicine*, *18*(3), 156–163.

https://doi.org/10.1016/j.molmed.2012.01.003

Peter, B., Raskind, W. H., Matsushita, M., Lisowski, M., Vu, T., Berninger, V. W., … Brkanac, Z. (2011). Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. *Journal of Neurodevelopmental Disorders*, *3*(1), 39–49. https://doi.org/10.1007/s11689-010-9065-0

Petkov, C. I., & Jarvis, E. D. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, *4*(AUG), 1–24. https://doi.org/10.3389/fnevo.2012.00012

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., … Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, *346*(6215), 1256846. https://doi.org/10.1126/science.1256846

Phillips, S., & Wilson, W. H. (2016). Commentary: Experimental evidence for compositional syntax in bird calls. *Frontiers in Psychology*, *7*(AUG), 1–7. https://doi.org/10.3389/fpsyg.2016.01171

Piaget, J. (1952). *Play, dreams and imitation in childhood*. Routledge.

Piazza, C., Cantiani, C., Akalin-Acar, Z., Miyakoshi, M., Benasich, A. A., Reni, G., … Makeig, S. (2016). ICA-derived cortical responses indexing rapid multi-feature auditory processing in six-month-old infants. *NeuroImage*, *133*, 75–87. https://doi.org/10.1016/j.neuroimage.2016.02.060

Pierce, L. J., Klein, D., Delcenserie, A., Genesee, F., Pierce, L. J., Klein, D., …

Genesee, F. (2015). Correction for Pierce et al., Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences*, *112*(8), E922. https://doi.org/10.1073/pnas.1501450112

Pika, S., & Bugnyar, T. (2011). The use of referential gestures in ravens (Corvus corax) in the wild. *Nature Communications*, *2*, 560. https://doi.org/10.1038/ncomms1567

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, *25*(5), 679–693. https://doi.org/10.1016/S0364-0213(01)00050-7

Poliak, S., Salomon, D., Elhanany, H., Sabanay, H., Kiernan, B., Pevny, L., … Peles, E. (2003). Juxtaparanodal clustering of Shaker-like K+ channels in myelinated axons depends on Caspr2 and TAG-1. *Journal of Cell Biology*, *162*(6), 1149–1160. https://doi.org/10.1083/jcb.200305018

Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., & Watwood, S. (2005). Animal behaviour: elephants are capable of vocal learning. *Nature*, *434*(7032), 455-456.

Popov, a V, Peresleni, a I., Ozerskii, P. V, Shchekanov, E. E., & Savvateeva-Popova, E. V. (2004). The fan-shaped and ellipsoid bodies of the brain central complex are involved in the control of courtship behavior and communicative sound production in Drosophila melanogaster males. *Rossiiskii Fiziologicheskii Zhurnal Imeni I.M. Sechenova / Rossiiskaia Akademiia Nauk*, *90*(4), 385–399.

Rafael E. Oliveras-Rentas, Lauren Kenworthy, Richard B. Roberson III, Alex Martin, and Wallace, G. L. (2013). WISC-IV profile in high-functioning autism spectrum disorders: impaired processing speed is associated with increased autism communication symptoms and decreased adaptive communication abilities. *Journal of Autism and Developmental Disorders,* 42(5), 655–664. http://doi.org/10.1007/s10803-011-1289-7.WISC-IV

Rattenborg, N. C., Martinez-Gonzalez, D., Roth, T. C., & Pravosudov, V. V. (2011). Hippocampal memory consolidation during sleep: A comparison of mammals and birds. *Biological Reviews*, *86*(3), 658–691. https://doi.org/10.1111/j.1469-185X.2010.00165.x

Rehkämper, G., Schuchmann, K. L., Schleicher, A., & Zilles, K. (1991). Encephalization in hummingbirds (Trochilidae). *Brain, behavior and evolution*, *37*(2), 85-91.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., … Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, *468*(7327), 1053–1060. https://doi.org/10.1038/nature09710

Rijksen, H. . (1978). A field study on Sumatran orangutans (Pongo pygmaeus abelii). (Doctoral dissertation, Veenman).

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. https://doi.org/10.1038/nn2072

Riva, V., Cantiani, C., Benasich, A. A., Molteni, M., Piazza, C., Giorda, R., … Marino,

C. (2017). From CNTNAP2 to Early Expressive Language in Infancy: The Mediation Role of Rapid Auditory Processing. *Cerebral Cortex*, 1–9. https://doi.org/10.1093/cercor/bhx115

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in neurosciences*, *21*(5), 188-194.

Rojas, D. C., Camou, S. L., Reite, M. L., & Rogers, S. J. (2005). Planum temporale volume in children and adolescents with autism. *Journal of Autism and Developmental Disorders*, *35*(4), 479–486. https://doi.org/10.1007/s10803-005-5038-7

Rousso, D. L., Pearson, C. A., Gaber, Z. B., Miquelajauregui, A., Li, S., Portera-Cailliau, C., … Novitch, B. G. (2012). Foxp-Mediated Suppression of N-Cadherin Regulates Neuroepithelial Character and Progenitor Maintenance in the CNS. *Neuron*, *74*(2), 314–330. https://doi.org/10.1016/j.neuron.2012.02.024

Roy, A. C., Curie, A., Nazir, T., Paulignan, Y., des Portes, V., Fourneret, P., & Deprez, V. (2013). Syntax at Hand: Common Syntactic Structures for Actions and Language. *PLoS ONE*, *8*(8), 1–11. https://doi.org/10.1371/journal.pone.0072677

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of cognitive neuroscience*, *26*(8), 1736-1747.

Scharff, C., & Petri, J. (2011). Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1574), 2124–2140.

https://doi.org/10.1098/rstb.2011.0001

Schreiweis, C., Bornschein, U., Burguière, E., Kerimoglu, C., Schreiter, S., Dannemann, M., … Graybiel, A. M. (2014). Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. *Proceedings of the National Academy of Sciences*, *111*(39), 14253–14258. https://doi.org/10.1073/pnas.1414542111

Schwartz, M. F., Faseyitan, O., Kim, J., & Coslett, H. B. (2012). The dorsal stream contribution to phonological retrieval in object naming. *Brain*, *135*(12), 3799–3814. https://doi.org/10.1093/brain/aws300

Shi, E. R., Zhang, E. Q. Martínez-Ferreiro, S. & Boeckx, C. 2015. Help Mr. X along, so say it with a beautiful song. Science of Aphasia 2015. Aveiro, Portugal.

Shi, L., Lin, Q., & Su, B. (2013). Human-specific hypomethylation of CENPJ, a key brain size regulator. *Molecular biology and evolution*, *31*(3), 594-604.

Shu, W., Cho, J. Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G. A., … Buxbaum, J. D. (2005). Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9643–8. https://doi.org/10.1073/pnas.0503739102

Shultz, S., & Vouloumanos, A. (2010). Three-Month-Olds Prefer Speech to Other Naturally Occurring Signals. *Language Learning and Development*, *6*(4), 241–257. https://doi.org/10.1080/15475440903507830

Simonyan, K. (2014). The laryngeal motor cortex: Its organization and connectivity. *Current Opinion in Neurobiology*, *28*, 15–21.

https://doi.org/10.1016/j.conb.2014.05.006

Singh, L., Liederman, J., Mierzejewski, R., & Barnes, J. (2011). Rapid reacquisition of native phoneme contrasts after disuse: You do not always lose what you do not use. *Developmental Science*, *14*(5), 949–959. https://doi.org/10.1111/j.1467-7687.2011.01044.x

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., … Held, R. M. (2014). Autism as a disorder of prediction. *Proceedings of the National Academy of Sciences,* 111(42), 15220–15225. http://doi.org/10.1073/pnas.1416797111

Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void: The need for multimodality in primate communication research. *Animal Behaviour*, *81*(5), 919–924. https://doi.org/10.1016/j.anbehav.2011.02.002

So, W.-C., Wong, M. K.-Y., Lui, M., & Yip, V. (2015). The development of co-speech gesture and its semantic integration with speech in 6- to 12-year-old children with autism spectrum disorders. *Autism*, *19*(8), 956–968. https://doi.org/10.1177/1362361314556783

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, *4*(3), 232–243. https://doi.org/10.1162/joc

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory Temporal Integration in Autism Spectrum Disorders. *The Journal of Neuroscience*, *34*(3), 691–697.

https://doi.org/10.1523/JNEUROSCI.3615-13.2014

Stiegler, L. N. (2015). Examining the Echolalia Literature: Where Do Speech-Language Pathologists Stand?. *American journal of speech-language pathology*, *24*(4), 750-762.

Stout, D., & Chaminade, T. (2012). Stone tools, language and the brain in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 75–87. https://doi.org/10.1098/rstb.2011.0099

Stout, D., & Chaminade, T. (2009). Making Tools and Making Sense: Complex, Intentional Behaviour in Human Evolution. *Cambridge Archaeological Journal*, *19*(1), 85. https://doi.org/10.1017/S0959774309000055

Strauss, K. A., Puffenberger, E. G., Huentelman, M. J., Gottlieb, S., Dobrin, S. E., Parod, J. M., … Morton, D. H. (2006). Recessive Symptomatic Focal Epilepsy and Mutant Contactin-Associated Protein-like 2. *New England Journal of Medicine*, *354*(13), 1370–1377. https://doi.org/10.1056/NEJMoa052773

Striedter, G. F. (2005). *Principles of Brain Evolution*. Sinauer, Sunderland, MA.

Striem-Amit, E., Almeida, J., Belledonne, M., Chen, Q., Fang, Y., Han, Z., … Bi, Y. (2016). Topographical functional connectivity patterns exist in the congenitally, prelingually deaf. *Scientific Reports*, *6*(February), 29375. https://doi.org/10.1038/srep29375

Subramanian, H. H., Balnave, R. J., & Holstege, G. (2008). The Midbrain Periaqueductal Gray Control of Respiration. *Journal of Neuroscience*, *28*(47), 12274–12283. https://doi.org/10.1523/JNEUROSCI.4168-08.2008

Subramanian, H. H., & Holstege, G. (2010). Periaqueductal gray control of breathing. In *New Frontiers in Respiratory Control. Advances in Experimental Medicine and Biology*, Homma I., Onimaru H., Fukuchi Y. (eds), vol 669. Springer, New York, NY.

Subramanian, H. H. (2013). Descending control of the respiratory neuronal network by the midbrain periaqueductal grey in the rat *in vivo*. *The Journal of Physiology*, *591*(1), 109–122. https://doi.org/10.1113/jphysiol.2012.245217

Taglialatela, J. P., Russell, J. L., Pope, S. M., Morton, T., Bogart, S., Reamer, L. A., … Hopkins, W. D. (2015). Multimodal communication in chimpanzees. *American Journal of Primatology*, *77*(11), 1143–1148. https://doi.org/10.1002/ajp.22449

Takeichi, M., & Abe, K. (2005). Synaptic contact dynamics controlled by cadherin and catenins. *Trends in Cell Biology*, *15*(4), 216–221. https://doi.org/10.1016/j.tcb.2005.02.002

Taylor, M. J., Charman, T., Robinson, E. B., Hayiou-Thomas, M. E., Happé, F., Dale, P. S., & Ronald, A. (2014). Language and traits of autism spectrum conditions: Evidence of limited phenotypic and etiological overlap. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *165*(7), 587-595.

Tavolga, W. N. (1956). Visual, chemical and sound stimuli as cues in the sex discriminatory behavior of the gobiid fish Bathygobius soporator. *Zoologica*, *41*(2), 49-64.

Tecumseh Fitch, W., & Reby, D. (2001). The descended larynx is not uniquely human.

Proceedings of the Royal Society B: Biological Sciences, *268*(1477), 1669–1675. https://doi.org/10.1098/rspb.2001.1704

Teichmann, M., Rosso, C., Martini, J. B., Bloch, I., Brugières, P., Duffau, H., … Bachoud-Lévi, A. C. (2015). A cortical-subcortical syntax pathway linking Broca's area and the striatum. *Human Brain Mapping*, *36*(6), 2270–2283. https://doi.org/10.1002/hbm.22769

Teki, S., Kumar, S., von Kriegstein, K., Stewart, L., Lyness, C. R., Moore, B. C. J., … Griffiths, T. D. (2012). Navigating the Auditory Scene: An Expert Role for the Hippocampus. *Journal of Neuroscience*, *32*(35), 12251–12257. https://doi.org/10.1523/JNEUROSCI.0082-12.2012

Teramitsu, I. (2004). Parallel FoxP1 and FoxP2 Expression in Songbird and Human Brain Predicts Functional Interaction. *Journal of Neuroscience*, *24*(13), 3152–3163. https://doi.org/10.1523/JNEUROSCI.5589-03.2004

Tesink, C. M. J. Y., Buitelaar, J. K., Petersson, K. M., Van Der Gaag, R. J., Kan, C. C., Tendolkar, I., & Hagoort, P. (2009). Neural correlates of pragmatic language comprehension in autism spectrum disorders. *Brain*, *132*(7), 1941–1952. https://doi.org/10.1093/brain/awp103

Tierney, A. L., & Nelson III, C. A. (2009). Brain development and the role of experience in the early years. *Zero to three*, *30*(2), 9.

Trevarthen, C. (2000). Musicality and the intrinsic motive pulse: evidence from human psychobiology and infant communication. *Musicae Scientiae*, (1999), 155–215. https://doi.org/10.1177/10298649000030S109

Truong, D. T., Rendall, A. R., Castelluccio, B. C., Eigsti, I.-M., & Fitch, R. H. (2015). Auditory Processing and Morphological Anomalies in Medial Geniculate Nucleus of Cntnap2 Mutant Mice. *Behavioral Neuroscience*, *129*(6), 731–43. https://doi.org/10.1037/bne0000096

Tsui, D., Vessey, J. P., Tomita, H., Kaplan, D. R., & Miller, F. D. (2013). FoxP2 Regulates Neurogenesis during Embryonic Cortical Development. *Journal of Neuroscience*, *33*(1), 244–258. https://doi.org/10.1523/JNEUROSCI.1665-12.2013

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory*. Tulving & W. Donaldson, (Eds.) https://doi.org/10.1017/S0140525X00047257

Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: role of the hippocampus. *Hippocampus*, *8*(3), 198–204. https://doi.org/10.1002/(sici)1098-1063(1998)8:3%3C198::aid-hipo2%3E3.0.co;2-g

Tyll, S., Budinger, E., & Noesselt, T. (2011). Thalamic influences on multisensory integration. *Communicative & integrative biology*, *4*(4), 378-381.

Uomini, N. T., & Meyer, G. F. (2013). Shared Brain Lateralization Patterns in Language and Acheulean Stone Tool Production: A Functional Transcranial Doppler Ultrasound Study. *PLoS ONE*, *8*(8), 1–9. https://doi.org/10.1371/journal.pone.0072693

Usui, N., Co, M., & Konopka, G. (2014). Decoding the molecular evolution of human cognition using comparative genomics. *Brain, Behavior and Evolution*, *84*(2), 103–116. https://doi.org/10.1159/000365182

Vail, A. L., Manica, A., & Bshary, R. (2013). Referential gestures in fish collaborative hunting. *Nature Communications*, *4*, 1765. https://doi.org/10.1038/ncomms2781

van der Lely, H. K. J., & Pinker, S. (2014). The biological basis of language: Insight from developmental grammatical impairments. *Trends in Cognitive Sciences*, *18*(11), 586–595. https://doi.org/10.1016/j.tics.2014.07.001

VanElzakker, M., Fevurly, R. D., Breindel, T., & Spencer, R. L. (2008). Environmental novelty is associated with a selective increase in Fos expression in the output elements of the hippocampal formation and the perirhinal cortex. *Learning & Memory*, *15*(12), 899–908. https://doi.org/10.1101/lm.1196508

Vanvuchelen, M., Roeyers, H., & De Weerdt, W. (2007). Nature of motor imitation problems in school-aged boys with autism. *Autism*, *11*(3), 225–240. https://doi.org/10.1177/1362361307076846

Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., … others. (2008). A functional genetic link between distinct developmental language disorders. *New England Journal of Medicine*, *359*(22), 2337–2345. http://www.nejm.org/doi/full/10.1056/NEJMoa0802828

Vernes, S. C., Oliver, P. L., Spiteri, E., Lockstone, H. E., Puliyadi, R., Taylor, J. M., … Fisher, S. E. (2011). FOXP2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genetics*, *7*(7). https://doi.org/10.1371/journal.pgen.1002145

Vivanti, G., & Hamilton, A. (2014). Imitation in autism spectrum disorders. In

Handbook of Autism and Pervasive Developmental Disorders, Vol 1: Diagnosis, Development, and Brain Mechanisms, ed. By Volkmar et al., 278–302. http://onlinelibrary.wiley.com/doi/10.1002/9781118911389.hautc12/full

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

Waldbillig, R. J. (1975). Attack, eating, drinking, and gnawing elicited by electrical stimulation of rat mesencephalon and pons. *Journal of comparative and physiological psychology*, *89*(3), 200.

Wallace, G. L., Anderson, M., & Happé, F. (2009). Brief report: Information processing speed is intact in autism but not correlated with measured intelligence. *Journal of Autism and Developmental Disorders*, *39*(5), 809–814. https://doi.org/10.1007/s10803-008-0684-1

Waller, B. M., & Micheletta, J. (2013). Facial Expression in Nonhuman Animals. *Emotion Review*, *5*(1), 54–59. https://doi.org/10.1177/1754073912451503

Wan, C. Y., Marchina, S., Norton, A., & Schlaug, G. (2012). Atypical hemispheric asymmetry in the arcuate fasciculus of completely nonverbal children with autism. *Annals of the New York Academy of Sciences*, *1252*(1), 332–337. https://doi.org/10.1111/j.1749-6632.2012.06446.x

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain*, *129*(4), 932–943. https://doi.org/10.1093/brain/awl032

Wang, R., Chen, C. C., Hara, E., Rivas, M. V., Roulhac, P. L., Howard, J. T., ... & Jarvis, E. D. (2015). Convergent differential regulation of SLIT-ROBO axon guidance genes in the brains of vocal learners. *Journal of Comparative Neurology*, *523*(6), 892-906.

Ward, B. J., Day, L. B., Wilkening, S. R., Wylie, D. R., Saucier, D. M., & Iwaniuk, a. N. (2012). Hummingbirds have a greatly enlarged hippocampal formation. *Biology Letters*, *8*(4), 657–659. https://doi.org/10.1098/rsbl.2011.1180

Warlaumont, A. S., & Finnegan, M. K. (2016). Learning to Produce Syllabic Speech Sounds via Reward-Modulated Neural Plasticity. *Plos One*, *11*(1), e0145096. https://doi.org/10.1371/journal.pone.0145096

Webb, D. M., & Zhang, J. (2005). FoxP2 in song-learning birds and vocal-learning mammals. *Journal of Heredity*, *96*(3), 212–216. https://doi.org/10.1093/jhered/esi025

Whiten, A., McGuigan, N., Marshall-Pescini, S., & Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2417–2428. https://doi.org/10.1098/rstb.2009.0069

Williams, H. (2001). Choreography of song, dance and beak movements in the zebra finch (Taeniopygia guttata). *The Journal of Experimental Biology*, *204*(Pt 20), 3497–506. https://doi.org/10.1002/(sici)1097-4695(19971105)33:5<602::aid-neu8>3.0.co

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, *9*(1), 11–29. https://doi.org/10.1007/BF01531288

Wohlgemuth, S., Adam, I., & Scharff, C. (2014). FoxP2 in songbirds. *Current Opinion in Neurobiology*, *28*, 86–93. https://doi.org/10.1016/j.conb.2014.06.009

Wray, C., Norbury, C. F., & Alcock, K. (2016). Gestural abilities of children with specific language impairment. *International Journal of Language & Communication Disorders*, *51*(2), 174–182. https://doi.org/10.1111/1460-6984.12196

Yamaguchi, A. (2001). Sex differences in vocal learning in birds. *Nature*, *411*(May), 257–258. https://doi.org/10.1038/35077143

Yeterian, E. H., & Pandya, D. N. (1998). of the Superior Temporal Region in Rhesus Monkeys. *Journal of Comparative Neurology*, *402*(May), 384–402.

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, *7*(6), 464–476. https://doi.org/10.1038/nrn1919

Zechner, D., Fujita, Y., Hülsken, J., Müller, T., Walther, I., Taketo, M. M., … Birchmeier, C. (2003). β-Catenin signals regulate cell growth and the balance between progenitor cell expansion and differentiation in the nervous system. *Developmental Biology*, *258*(2), 406–418. https://doi.org/10.1016/S0012-1606(03)00123-4

Zhang, J. (2003). Evolution of the Human ASPM Gene, a Major Determinant of Brain Size, *2070*(December), 2063–2070.

Zhang, J., Webb, D. M., & Podlaha, O. (2002). Accelerated protein evolution and origins of human-specific features: FOXP2 as an example. *Genetics*, *162*(4), 1825–1835. https://doi.org/10.1038/35102048